

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:32:13 ; Search time 2266.09 Seconds
(without alignments)
291.578 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagcgttctaagggccgc 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

59: gb_est33:*

60: gb_est34:*

61: gb_est35:*

62: gb_est36:*

63: gb_est37:*

64: gb_est38:*

65: em_est27:*

66: em_est28:*

67: em_est29:*

68: em_est30:*

69: gb_est39:*

70: gb_est40:*

71: gb_est41:*

72: gb_est42:*

73: gb_est43:*

74: gb_est44:*

75: em_est31:*

76: em_est32:*

77: em_est33:*

78: em_est34:*

79: gb_gss1:*

80: gb_gss2:*

81: gb_gss3:*

82: gb_gss4:*

83: em_gss1:*

84: em_gss2:*

85: em_gss3:*

86: em_gss4:*

87: gb_gss5:*

88: gb_gss6:*

89: gb_gss7:*

90: gb_gss8:*

91: gb_gss9:*

92: em_gss5:*

93: em_gss6:*

94: em_gss7:*

95: em_gss8:*

96: em_gss9:*

97: em_gss10:*

98: em_gss11:*

99: gb_gss10:*

100: gb_gss11:*

101: em_gss12:*

102: gb_gss12:*

103: gb_gss13:*

104: gb_gss14:*

105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	8		DB ID	Description
			Length	Length		
1	34.2	19.5	203	24	H84039	H84039 yv88e11.s1
2	33.8	19.3	413	28	AA085404	AA085404 zn07f02.s

www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 434 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

Location/Qualifiers
1. .324

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:909465"
/clone_lib="NCI_CGAP_Alvi"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"

/note="Vector: pAMP10; mRNA made from alveolar rhabdomyosarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 78 a 79 c 79 g 88 t

ORIGIN

Query Match 19.0%; Score 33.2; DB 34; Length 324;
Best Local Similarity 50.6%; Pred. No. 4.5;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4 ttcccaccatgcatcagatataaaagcaaatctctaaatttatagtgataaactgaactag 63

DB 21 TGCACACCATGGCTAAGGTCACAACTTTAGAAAAATAAAATTCATGTAAATCCCAAGAT 80

QY 64 gaggttgccatggttcagaaaccatgatcattcatcgtgtaaacaccattcgcggttgc 123

DB 81 AGGGCCCCCTGGTGGTTTCATCCCTCTCTCTTTGTGAGCCCAAGTCCTCTCTTTTCT 140

QY 124 acggaagtgccttcattatcacagaggaagccg 161

DB 141 TGGGAATGGAGCCCGCAGAGCCACAGAGGAAGCAG 178

RESULT

AA931139 5
LOCUS AA931139 327 bp mRNA EST 07-JUL-1998
DEFINITION O070b09.s1 NCI_CGAP_G4 Homo sapiens cDNA clone IMAGE:1571513 3', mRNA sequence.

ACCESSION AA931139

VERSION AA931139.1 GI:3085525

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jan 9, 1998 this sequence version replaced gi:716604.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

Emmert-Buck, M.D., Ph.D.

cDNA Library prepared by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbrrp/image/image.html

FEATURES

Location/Qualifiers
1. .327

Insert Length: 458 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 314.

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1571513"

/clone_lib="NCI_CGAP_G4"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 81 a 81 c 79 g 86 t

ORIGIN

Query Match 19.0%; Score 33.2; DB 40; Length 327;

Best Local Similarity 50.6%; Pred. No. 4.5;

Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4 ttcccaccatgcatcagatataaaagcaaatctctaaatttatagtgataaactgaactag 63

DB 26 TGCACACCATGGCTAAGGTCACAACTTTAGAAAAATAAAATTCATGTAAATCCCAAGAT 85

QY 64 gaggttgccatggttcagaaaccatgatcattcatcgtgtaaacaccattcgcggttgc 123

DB 86 AGGGCCCCCTGGTGGTTTCATCCCTCTCTCTTTGTGAGCCCAAGTCCTCTCTTTTCT 145

QY 124 acggaagtgccttcattatcacagaggaagccg 161

DB 146 TGGGAATGGAGCCCGCAGAGCCACAGAGGAAGCAG 183

RESULT

AA524847 6
LOCUS AA524847 339 bp mRNA EST 20-AUG-1997
DEFINITION ng46e02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:937850 3', mRNA sequence.

ACCESSION AA524847

VERSION AA524847.1 GI:2265775

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On May 5, 1995 this sequence version replaced gi:798189.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,

M.D., Ph.D.

cDNA Library preparation: M. Bento Soares, Ph.D.

DNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbrrp/image/image.html

FEATURES

Location/Qualifiers
1. .339

Insert Length: 429 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:937850"

/clone_lib="NCI_CGAP_Co3"

/sex="pooled"	
/tissue_type="colon"	
/lab_host="DH108"	
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."	
BASE COUNT	86 a 85 c 87 g 81 t
ORIGIN	
Query Match	19.0%; Score 33.2; DB 34; Length 339;
Best Local Similarity	50.6%; Pred. No. 4.5;
Matches	80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY	4 ttccaccatggtacgtatataaaagcaattctaaatttaggtataactgaactag 63
Db	19 TGAACACCATGGCTAAGGTCACAACTGTAGAAAAATAAATTCATGTAAATCCCAAGAT 78
QY	64 gagggtccatggttcagaaccatgatacttcattcgtgttaaacattcgggggtgtc 123
Db	79 AGGCCCTTGGGTTTCATCCCTCTCTTCGTGAGCCCAAGCTCTCTTTTCCT 138
QY	124 acggaagtgccttgctattacacagagagaaagccg 161
Db	139 TGGGATGGAGCCGGCAGAGGCCACAGAGGAAGCAG 176
RESULT	7
H77931	
LOCUS	H77931 419 bp mRNA EST 09-NOV-1995
DEFINITION	ys10404.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:214375 3', mRNA sequence.
ACCESSION	H77931
VERSION	H77931.1
KEYWORDS	GI:1056020
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 419) Hillier L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	On Apr 14, 1993 this sequence version replaced gi:692549. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 719 High quality sequence stops: 329 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 719 Std Error: 0.00 Seq primer: Promeg -21ml3 High quality sequence stop: 329. Location/Qualifiers 1. .419 /organism="Homo sapiens"
FEATURES	
source	


```

Insert Length: 1202   Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence shot: 243.
Location/Qualifiers
1. 273
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1913078"
/clone_lib="NCI_CGAP_Co8"
/tissue_type="adenocarcinoma"
/lab_host="PH103"
/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(df) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
68 a 70 c 56 g 69 t
BASE COUNT

```

Search completed: April 8, 2000, 01:04:00
Job time: 5507 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:34:45 ; Search time 371.99 Seconds
(without alignments)
-1428.445 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatggtacg.....aagcgttctaagcgccgc 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_on.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: gb_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_on.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48.2	27.5	588	5	A73353	A73353 Sequence 3
2	48.2	27.5	588	5	AR013794	AR013794 Sequence
3	48.2	27.5	588	5	AR047910	AR047910 Sequence
4	48.2	27.5	588	5	I13096	I13096 Sequence 3
5	48.2	27.5	588	5	I25469	I25469 Sequence 3
6	48.2	27.5	591	5	A01422	A01422 mRNA of lam
7	45	25.7	1791	1	CLOTETOX	M12739 C.tetani te
8	45	25.7	1851	5	E01135	E01135 cDNA encodi
9	45	25.7	4338	1	CTTETOXG	X06214 Clostridium
10	45	25.7	4400	1	CTTETANG	X04436 Clostridium
11	41.8	23.9	390	5	E08086	E08086 cDNA encodi
12	41.8	23.9	390	5	E08087	E08087 cDNA encodi
13	41.8	23.9	390	5	E08088	E08088 cDNA encodi
14	41.8	23.9	390	5	E08089	E08089 Modified al
15	41.8	23.9	390	5	E08091	E08091 Modified al
16	41.8	23.9	390	24	E09003	E09003 DNA encodin
17	41.8	23.9	393	5	E06573	E06573 cDNA encodi
18	41.8	23.9	483	34	DEPDER1	D10447 Dermatophag
19	41.8	23.9	485	34	DEPDER3	D10448 Dermatophag
20	41.8	23.9	513	5	AR025156	AR025156 Sequence
21	41.8	23.9	513	5	E03185	E03185 DNA encodin
22	41.8	23.9	516	5	AR025155	AR025155 Sequence
23	41.8	23.9	517	5	E03184	E03184 DNA encodin
24	41.8	23.9	517	34	S70378	S70378 Der f II-gr
25	41.8	23.9	528	34	DEPDER2	D10448 Dermatophag
26	40.2	23.0	491	5	A73357	A73357 Sequence 7
27	40.2	23.0	491	5	AR013796	AR013796 Sequence
28	40.2	23.0	491	5	AR047912	AR047912 Sequence
29	40.2	23.0	491	5	I13098	I13098 Sequence 7
30	40.2	23.0	491	5	I25471	I25471 Sequence 7
31	37	21.1	390	5	E08090	E08090 Modified al
32	33.6	19.2	503	35	AF047613	AF047613 Euroglyph
33	33.6	19.2	505	35	AF047614	AF047614 Euroglyph
34	33.4	19.1	2289	34	CELCPR6G	L39939 Caenorhabdi
35	33.4	19.1	34514	35	CELC25B8	U41556 Caenorhabdi
36	33.4	19.1	65716	11	AC003657	AC003657 Homo sapi
37	33.4	19.1	161814	44	AC016676	AC016676 Homo sapi
38	33.4	19.1	172316	44	AC016543	AC016543 Homo sapi
39	32.8	18.7	95238	45	AC017105	AC017105 Homo sapi
40	32.8	18.7	167561	45	AC016923	AC016923 Homo sapi
41	32.4	18.5	3780	34	CEJUNCMECB	L02878 Caenorhabdi
42	32.2	18.4	30332	34	CEC43C3	Z47067 Caenorhabdi
43	32	18.3	34751	34	CEF26D2	Z81513 Caenorhabdi
44	31.8	18.2	1772	7	MTNGCYTBC	X87997 M.galapoda
45	31.8	18.2	186598	42	AC006334	AC006334 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS A73353 A73353 588 bp DNA
DEFINITION Sequence 3 from Patent WO9424281.
ACCESSION A73353
VERSION A73353.1 GI:6064126
KEYWORDS

PAT 15-OCT-1999

```

ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 588)
AUTHORS       Garman,R.D., Greenstein,J.L., Kuo,M., Rogers,B.L., Franzen,H.M.,
              Chen,X., Evans,S. and Shaked,Z.
TITLE        T cell epitopes of the major allergens from dermatophagoides (house
              dust mite)
JOURNAL       Patent: US 5820862-A 3 13-OCT-1998;
FEATURES      Location/Qualifiers
              .. 1..588
              /organism="unknown"
BASE COUNT    206 a 105 c 98 g 179 t
ORIGIN
Query Match   27.5%; Score 48.2; DB 5; Length 588;
Best Local Similarity 94.3%; Pred. No. 0.00031;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0; 0; 0;

Qy 61 taggaggttgccattggttcagaaccatgtatcatcttcctggttaaacattc 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 TACCAGGATGCGATGGTTTCAGAACCATGTATCATTCATCGTGGTAAACCATTC 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT	4			
I13096				
LOCUS	I13096	588 bp	DNA	
DEFINITION	Sequence 3 from patent US 5433948.			
			PAT	26-JUL-1995

REFERENCE	1 (bases 1 to 588)
AUTHORS	Thomas, W. R. and Chua, K.
TITLE	Cloning and sequencing of allergens of dermatophagoides (house dust mite)
JOURNAL	Patent: US 5433948-A 3 18-JUL-1995;
FEATURES	location/Qualifiers
source	1. 588
	/organism="unknown"
BASE COUNT	206 a 105 c 98 g 179 t
ORIGIN	

	Query Match	27.5%	Score 48.2	DB 5	Length 588
	Best Local Similarity	94.3%	Pred. No. 0.00031		
	Matches 50	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Qy	61	tagaggtg	gcattggttcagagaccatgtatcattcatcgtggttaaacattc	113	
Db	172	TACAGAGATGCATGGTT	CAGAACCATGTATCATTCATCGTGGTAAACCATTC	224	

REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 588)
TITLE	Thomas,W.R. and Chua,K. Cloning and sequencing of allergens of dermatophagoides (house dust mite)
JOURNAL	Patent: US 552142-A 3 03-SEP-1996;
FEATURES	Location/Qualifiers
source	1. .588 /organism="unknown"

MEDLINE	6085672
COMMENT	Draft entry and sequence in computer readable form for [1] kindly provided by N.F.Fairweather, 05-AUG-1986. The last 122 amino acids of fragment B are found at position 1 to 366. Fragment C of the tetanus toxin is located at positions 367-1719.
FEATURES	Location/Qualifiers source 1..1791 /organism="Clostridium tetani" /db_xref="taxon:1513" CDS 1..1722 /note="tetanus toxin" /codon_start=1 /transl_table=11 /protein_id="AA23282.1" /db_xref="GI:144921" /translation="RSLEYQVDAIKIIDIYEKIIYSGPKQEIQADIENLNKKLEKA NKAMINIFMRESSRFLVNOMINEAKQLLEFFDQSKNLMQYIKANSFGITGL KKEISKINKVFSTPIPFYSKNLDCWDNEEDIDVLKSTILNDINDIIDISLVG NSSVITYPDALVPNGINGKAIHLVNNESSEVIHVKAMDIEYDNMFNFVSFWLRPFP VASHEOYGTNEYSIISMKKHSLSIGSWSLKGNNLIWTKDSAGEVRQITFRD LPDKFNAYLANKWFTITNDRLSSANLYINGVMGSASITGLGAIRENNITLKLD R CANNQYVSIDAFRECKALNKPEKLTYSLSITFLDFGNQRIODETYELIPVA SSSKDQLKNIIDYIMLTNPATYTNGKLIYYRRLYNGLFKFIKRYPNNEIDSFVKS GDFIKLYSYNNNEHIIVGYPDGNAPFNNDRLIRLVGYNAGPIPLKYKMEAVKLRLDKT YSVOLKLYDDKNASLGVLGVTHNGQIGNDPNRDILLIASNNYFHNHDKILGCDWYFYPT DSGWTND"
BASE COUNT	727 A 197 C 276 G 591 T
ORIGIN	1 bp upstream of BgIII site.
Query Match	25.7%; Score 45; DB 1; Length 1791;
Best Local Similarity	100.0%; Pred. No. 0.0027;
Matches	45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	18 cagtataaaagcaattctaatatagggtataactgaacta 62
Db	262 CAGTATAAAGCAAAATTCTAAATTTATAGGTATAACTGAAC TA 306
RESULT	8
E01135	
LOCUS	E01135 1851 bp DNA PAT 29-SEP-1997
DEFINITION	CDNA encoding tetanustoxin C fragment.
ACCESSION	E01135
VERSION	E01135.1 GI:2169394
KEYWORDS	JP 1987051994-A/1; Clostridium tetani.
SOURCE	Clostridium tetani
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE	1 (bases 1 to 1851)
AUTHORS	Neiruf,F.
TITLE	CLONED ANTIGEN
JOURNAL	Patent: JP 1987051994-A 1 06-MAR-1987; WELLCOME FOUND LTD:THE
COMMENT	OS Clostridium tetani PN JP 1987051994-A/1 PD 06-MAR-1987 PF 27-JUN-1986 JP 1986149849 PR 28-JUN-1985 GB 85 8516442 PI NEIRU FUREIZAA FUEAAUEZAA PC C12N15/00,A61K39/08,C07K13/00,C12P21/02,(C12N15/00,C12R1:145); (C12P21/02, PC C12R1:19); CC strandedness: Double; CC topology: Linear; CC hypothetical: No; CC anti-sense: No; CC *source: strain-CN 3911; CC *source: clone-pTET 12; FH Key Location/Qualifiers PH

BASE COUNT	206 A 105 C 98 G 179 T
ORIGIN	
Query Match	27.5%; Score 48.2; DB 5; Length 588;
Best Local Similarity	94.3%; Pred. No. 0.00031;
Matches	50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	61 taggagtgccatgggttcagaaccatgatcatcctcggtgtaaacattc 113
Db	172 TACCAGATGCCATGGTTCAGAACCATGATCATTCATCGTGTAACCATTC 224
RESULT	6
A01422	
LOCUS	A01422 591 bp mRNA PAT 10-MAR-1993
DEFINITION	mRNA of lambda gt11 pII(Cl).
ACCESSION	A01422
VERSION	A01422.1 GI:344350
KEYWORDS	. synthetic construct. artificial sequence.
SOURCE	1 (bases 1 to 591)
ORGANISM	Cloning of MITE ALLERGENS
REFERENCE	Patent: WO 8810297-A 3 29-DEC-1988;
AUTHORS	Journal
TITLE	Location/Qualifiers
JOURNAL	1..591
FEATURES	/organism="synthetic construct" /db_xref="taxon:32630" 72..509 /codon_start=1 /transl_table=11 /product="Der p II" /protein_id="CAA00161.1" /db_xref="GI:344351" mat_peptide 120..506 /product="Der p II"
BASE COUNT	206 A 107 C 98 G 180 T
ORIGIN	
Query Match	27.5%; Score 48.2; DB 5; Length 591;
Best Local Similarity	94.3%; Pred. No. 0.00031;
Matches	50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	61 taggagtgccatgggttcagaaccatgatcatcctcggtgtaaacattc 113
Db	172 TACCAGATGCCATGGTTCAGAACCATGATCATTCATCGTGTAACCATTC 224
RESULT	7
CLOTETTOX	
LOCUS	CLOTETTOX 1791 bp DNA BCT 26-APR-1993
DEFINITION	C.tetani tetanus toxin gene.
ACCESSION	M12739
VERSION	M12739.1 GI:144920
KEYWORDS	tetanus toxin. C.tetani (strain CN3911, derivative of Harvard strain) DNA, clones pTet[1.8].
SOURCE	Clostridium tetani
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE	1 (bases 1 to 1791)
AUTHORS	Fairweather,N.F., Lyness,V.A., Pickard,D.J., Allen,G.P. and Thomson,R.O.
TITLE	Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli
JOURNAL	J. Bacteriol. 165, 21-27 (1986)


```

FT CDS >1..1722 /products-'tetanustoxin'
FT mat_peptide >1..1719
FT 3'UTR /product-'tetanustoxin'
FT mRNA >1..1851
FT FEATURES Location/Qualifiers
   source 1..1851
            /organism="Clostridium tetani"
            /db_xref="taxon:1513"
BASE COUNT 753 a 204 c 279 g 615 t
ORIGIN
Qy 18 cagtatataaaagcaaatctctaaatttaggtataactgaacta 62
    |||||||
Db 262 CAGTATATAAAAGCAAAATCTAAATTTATAGGTATAACTGAAC 306
    |||||||

Query Match 25.7%; Score 45; DB 5; Length 1851;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 cagtatataaaagcaaatctctaaatttaggtataactgaacta 62
    |||||||
Db 262 CAGTATATAAAAGCAAAATCTAAATTTATAGGTATAACTGAAC 306
    |||||||

RESULT 9
CITEOXG 4338 bp DNA BCT 12-SEP-1993
LOCUS Clostridium tetani gene for tetanus toxin.
DEFINITION X06214.1 GI:40773
ACCESSION X06214.1
VERSION X06214.1
KEYWORDS tetanus toxin.
SOURCE Clostridium tetani.
ORGANISM Clostridium tetani.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4338)
AUTHORS Fairweather, N.F. and Lyness, V.A.
TITLE The complete nucleotide sequence of tetanus toxin
JOURNAL Nucleic Acids Res. 14 (19), 7809-7812 (1986)
MEDLINE 87040747
COMMENT Data kindly reviewed (25-Nov-1988) by Lyness V.A.
FEATURES
   source 1..4338
            /organism="Clostridium tetani"
            /strain="CN3911"
            /db_xref="taxon:1513"
            /clone="pAT153"
   CDS 322..4269
        /note="tetanus toxin (AA 1-1315)"
        /codon_start=1
        /transl_table=11
        /protein_id="CAA29564.1"
        /db_xref="GI:40774"
        /db_xref="SWISS-PROT:P04958"
        /translation="MPITINFRYSDPVNNTIMMEPPYCKGLDIYKAFKITDRIW
        IVERYEGTKPEDNPSSLEGASEYDNPYLRDSDKDRFLQTMVKLFNRIKNV
        AGEALLDILKINAIPLVGNISLLDGFDTNSVSNLLEQDPGSGATKSAMLNLIIF
        GPGVLNKEVRGIVLRVDKNYPCRDGFGSIMQAFCEPVYFDNVENITSLTI
        GSKYFQDPALLMHELIVHLGLYGMQVSSHEIIPSKQEIYMQHTYPISAELFTFG
        GDANLISIDIKNDLYEKTNDYKAANKLSQVTSNDPNIDIDSYKIOOQYQFDK
        DSNQGIYVNEKDFQLYNSIMYGFTEIELGKFNITRLSYFSMNDVPKIPNLLDDT
        IYNDTEGNIESKDLSEYKQNMVRNTNFRNVDGSLVSKLIGLCKIKIIPPTNIRE
        NLYNTASVLDLGGELCIKINEDLTFIAEKNSRSEEPFQDEIVSYNTKKNLNFNS
        LDKLIVYLNQSKITLPNDRTPTKGIPIAPYKSNAASTIEIHNIDNTIYQVLYA
        QKSPITLQRTMNSVDALINISYIFPSYISKVQAGQILFLQVRDIIIDDT
        NNESSQKTTIDKISDVSTIVPIGPNALIVKQEGNFTGALETTGVLLLEIPEYITL
        PVRAALSASTQKEIKITIDNPLEKRYEKWIEYKLVKAKWLGTVTQPKRSYQ
        MYRSEYQVDAIKIIDEYKIIYSGPDKQIADENLNKLNKEKANKAMINIFMR
        ESSRSEYQVDAIKIIDEYKIIYSGPDKQIADENLNKLNKEKANKAMINIFMR
        TPPIPSYKNDLCHVDNEDIDVILKSTILNLDINDIISDISGNSSVIYPAQL
        VPGNGKALHVNESSEVIVHKAMDIEINMNFNTVSWLRVFPVPSASHLEQGTGN
        EYIISLSSKHSLSIGSGVSLKGNLNIWLTLSKAGEVQRITFDLPDKFNAYLANK
        WVFITIDNRLSSANLYNGVLMSGAEITGLGAIREDNNITLKLDRCNNNOYVSDIK
        FRIFCKALNPKEIEKLYTSITFLRDFWGNPLRYDTEYIYLPVASSSKDQVLKNIT
        DYMILTNPASVTNGKLNIIYRRLYNGLFIKRIKRYTPNNEIDSVKSGDFIKLYSVNN
        NEHIVYKPDGNAFNENLDRILRVGYNAFCIPLYKKRKAVALKRLKLTYSYQLKLYDDKN
        ASLGLVGTGHNQIGNDNRDILLIAGNWFNHLKDKILGDMYFVPTDEGWIND"
BASE COUNT 1783 a 522 c 640 g 1393 t
ORIGIN
Query Match 25.7%; Score 45; DB 1; Length 4338;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 cagtatataaaagcaaatctctaaatttaggtataactgaacta 62
    |||||||
Db 2809 CAGTATATAAAAGCAAAATCTAAATTTATAGGTATAACTGAAC 2853
    |||||||

RESULT 10
CITEOXG 4400 bp DNA BCT 12-JUL-1995
LOCUS Clostridium tetani gene for tetanus toxin.
DEFINITION X04436
ACCESSION X04436
VERSION X04436.1 GI:40769
KEYWORDS neurotoxin; plasmid; tetanus toxin; toxin.
SOURCE Clostridium tetani.
ORGANISM Clostridium tetani.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4400)
AUTHORS Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J.,
Weller, U., Hudel, M., Habermann, E. and Niemann, H.
TITLE Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins
JOURNAL EMBO J. 5 (10), 2495-2502 (1986)
MEDLINE 87053814
COMMENT See also Fairweather et al. (1986) J.Bacteriol. 165, 21-27. Data
kindly reviewed (12-MAY-1987) by U.Eisel.
FEATURES
   source 1..4400
            /organism="Clostridium tetani"
            /db_xref="taxon:1513"
            /note="open reading frame (16AA) (1 is 2nd base in codon)"
            /codon_start=3
            /transl_table=11
            /protein_id="CAA28032.1"
            /db_xref="GI:809667"
            /db_xref="SPTREMBL:Q46456"
            /translation="ALKKLEPIVKNLNI"
            /note="region of inverted repeats"
   misc_feature 37..250
        /note="tetanus toxin precursor (AA 1-1315)"
   RBS 267..272
        /note="rRNA binding site"
   mat_peptide 281..1651
        /product="tetanus toxin L-chain (AA 1-457)"
        /db_xref="GI:40770"
   CDS 281..4228
        /codon_start=1
        /transl_table=11
        /product="tetanus toxin precursor (AA 1-1315)"
        /protein_id="CAA28033.1"
        /db_xref="GI:40770"
        /db_xref="SWISS-PROT:P04958"
        /translation="MPITINFRYSDPVNNTIMMEPPYCKGLDIYKAFKITDRIW
        IVERYEGTKPEDNPSSLEGASEYDNPYLRDSDKDRFLQTMVKLFNRIKNV
        AGEALLDILKINAIPLVGNISLLDGFDTNSVSNLLEQDPGSGATKSAMLNLIIF
        GPGVLNKEVRGIVLRVDKNYPCRDGFGSIMQAFCEPVYFDNVENITSLTI
        GSKYFQDPALLMHELIVHLGLYGMQVSSHEIIPSKQEIYMQHTYPISAELFTFG
        GDANLISIDIKNDLYEKTNDYKAANKLSQVTSNDPNIDIDSYKIOOQYQFDK
        DSNQGIYVNEKDFQLYNSIMYGFTEIELGKFNITRLSYFSMNDVPKIPNLLDDT
        IYNDTEGNIESKDLSEYKQNMVRNTNFRNVDGSLVSKLIGLCKIKIIPPTNIRE
        NLYNTASVLDLGGELCIKINEDLTFIAEKNSRSEEPFQDEIVSYNTKKNLNFNS
        LDKLIVYLNQSKITLPNDRTPTKGIPIAPYKSNAASTIEIHNIDNTIYQVLYA
        QKSPITLQRTMNSVDALINISYIFPSYISKVQAGQILFLQVRDIIIDDT
        NNESSQKTTIDKISDVSTIVPIGPNALIVKQEGNFTGALETTGVLLLEIPEYITL
        PVRAALSASTQKEIKITIDNPLEKRYEKWIEYKLVKAKWLGTVTQPKRSYQ
        MYRSEYQVDAIKIIDEYKIIYSGPDKQIADENLNKLNKEKANKAMINIFMR
        ESSRSEYQVDAIKIIDEYKIIYSGPDKQIADENLNKLNKEKANKAMINIFMR
        TPPIPSYKNDLCHVDNEDIDVILKSTILNLDINDIISDISGNSSVIYPAQL
        VPGNGKALHVNESSEVIVHKAMDIEINMNFNTVSWLRVFPVPSASHLEQGTGN
        EYIISLSSKHSLSIGSGVSLKGNLNIWLTLSKAGEVQRITFDLPDKFNAYLANK
        WVFITIDNRLSSANLYNGVLMSGAEITGLGAIREDNNITLKLDRCNNNOYVSDIK
        FRIFCKALNPKEIEKLYTSITFLRDFWGNPLRYDTEYIYLPVASSSKDQVLKNIT

```

PVTAALSIAESSTOKELIKITIDNLEKRYEKKWIEVLYKAKWLGTVNTOPQRSVQ
 MRSLEYQDAIKIIDIYKIEYKISGPKDEQIADEINLNKLEKANKAMININIFMR
 ESRSLVNMINEAKKQLEFDTQSKNLMQYKANSKFIETELKLESKINKVFS
 TTPFSYKNDLQWVNEEDIDVILKSTILNDINNDIISDGSNSVITYPDQOL
 VPGINGKALHLVNNESSEVIVHKAMDIENMFNFTVFWLRVPKVVSASHLQYGN
 EYKISGSMKKHSLISGGSVSLKGNLWTLKDSAGEVRQITFDLPDKFNAYLANK
 WYFITNDRLSSANLYINGLMSAEITGLGAIREDNNTILKIDRCNNNNQVYSIDK
 FRIFCALNPKEIEKIYTSVLSITFLDFWGNELRIDTEYILIPVASSSKDVQKNT
 DMYLNPASYNGKLNIIYRRLYNGLFKFIKRYTPNNEIDSFVKSGDFIKLYSVNN
 NHHVGYPKDGNFNLDRILRYGYNAPGLPKMEAVKRLDKTYSVOLKLYDDKN
 ASLGLVGTNGDGNPNRDLILASNNWFNHLKDKILGCDWYFVPTDEGTND"
 1652. .4225
 mat_peptide /products="tetanus toxin H-chain (AA 458-1315)"
 misc_feature /note="region of inverted repeats"
 468. .4394
 BASE COUNT 1808 a 532 c 644 g 1416 t
 ORIGIN

Query Match 25.7%; Score 45; DB 1; Length 4400;
 Best Local Similarity 100.0%; Pred. No. 0.0028;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 cagtataaaacaaattctaaatttataggtataactgaacta 62
 |||||
 Db 2768 CAGTATATAAAGCAATCTAAATTTATAGGTATACTGAACTA 2812

RESULT 11
 E08086 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 ACCESSION E08086
 VERSION E08086.1 GI:2176211
 KEYWORDS JP 1994253851-A/10.
 SOURCE Dermatophagoides farinae.
 ORGANISM Dermatophagoides farinae
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcotiformes; Astigmata; Pyroglyphoidea;
 Pyroglyphidae; Dermatophagoides.
 1 (bases 1 to 390)
 Takai,T., Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
 MODIFIED MAJOR MITE ALLERGEN AND ITS PRODUCTION
 Patent: JP 1994253851-A 10 13-SEP-1994;
 ASAHI BREWERIES LTD, TORII YAKUJIN KK, NIKKA UISUKII KK
 OS Dermatophagoides farinae (acarid)
 PN JP 1994253851-A/10
 PD 13-SEP-1994
 PF 04-MAR-1993 JP 1993139793
 PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA
 HIROSHI,
 PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
 (C12P21/02);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT source 1. .390
 FT /organism="Dermatophagoides farinae".

BASE COUNT 125 a 78 c 76 g 111 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaacacatgatacattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATCGTGTCTGATCCATCATCCATCGTGTGTAACCATTC 105

RESULT 13
 E08088 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 ACCESSION E08088
 VERSION E08088.1 GI:2176213
 KEYWORDS JP 1994253851-A/12.
 SOURCE Dermatophagoides farinae.
 ORGANISM Dermatophagoides farinae

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaacacatgatacattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATCGTGTCTGATCCATCATCCATCGTGTGTAACCATTC 105

RESULT 12
 E08087 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 ACCESSION E08087
 VERSION E08087.1 GI:2176212
 KEYWORDS JP 1994253851-A/11.
 SOURCE Dermatophagoides farinae.
 ORGANISM Dermatophagoides farinae
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcotiformes; Astigmata; Pyroglyphoidea;
 Pyroglyphidae; Dermatophagoides.
 1 (bases 1 to 390)
 Takai,T., Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
 MODIFIED MAJOR MITE ALLERGEN AND ITS PRODUCTION
 Patent: JP 1994253851-A 11 13-SEP-1994;
 ASAHI BREWERIES LTD, TORII YAKUJIN KK, NIKKA UISUKII KK
 OS Dermatophagoides farinae (acarid)
 PN JP 1994253851-A/11
 PD 13-SEP-1994
 PF 04-MAR-1993 JP 1993139793
 PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA
 HIROSHI,
 PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
 (C12P21/02);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT source 1. .390
 FT /organism="Dermatophagoides farinae".

BASE COUNT 126 a 77 c 75 g 112 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaacacatgatacattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATCGTGTCTGATCCATCATCCATCGTGTGTAACCATTC 105

RESULT 13
 E08088 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 ACCESSION E08088
 VERSION E08088.1 GI:2176213
 KEYWORDS JP 1994253851-A/12.
 SOURCE Dermatophagoides farinae.
 ORGANISM Dermatophagoides farinae

```

FT             source      1. 390
FT             /organism='Artificial sequences'
FEATURES
source      Location/Qualifiers
1. 390
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT    127 a 79 c 76 g 108 t
ORIGIN

Query Match      23.9%; Score 41.8; DB 5; Length 390;
Best Local Similarity 86.8%; Pred. NO. 0.023;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttccatggttcagaaccatgtatcattcatcgtggttaaccattc 113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 TCGATGTTGCCATGGTTCTGATCCATGCATCCATCCATCGTGGTAAACCATTC 105

RESULT 15
E08091
LOCUS      E08091      390 bp      DNA      PAT      29-SEP-1997
DEFINITION Modified allergen DNA of Dermatophagoides farinae.
ACCESSION E08091
VERSION   E08091.1 GI:2176216
KEYWORDS  JP 1994253851-A/15.
SOURCE    unidentified.
ORGANISM  unidentified
           unclassified.

REFERENCE   1 (bases 1 to 390)
AUTHORS   Takai,T.,Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
TITLE     MODIFIED MAJOR MITE ALLERGEN AND LTS PRODUCTION
JOURNAL   Patent:JP 1994253851-A 15 13-SEP-1994;
          ASAHI BREWERIES LTD, TORII YAKUHI KK, NIKKA UISUKII KK
COMMENT   OS None
          OC Artificial sequences.
          PN JP 1994253851-A/15
          PD 13-SEP-1994
          PF 04-MAR-1993 JP 1993139793
          PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA HIROSHI,
          ANDO TORU, HIRAI MITSUO
          PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
          (C12P21/02)
          CC C12R1:19;
          CC strandedness: Double;
          CC topology: Linear;
          FH Key
          PH Location/Qualifiers
          FT source      1. 390
          FT             /organism='Artificial sequences'
FEATURES
source      Location/Qualifiers
1. 390
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT    127 a 78 c 76 g 109 t
ORIGIN

Query Match      23.9%; Score 41.8; DB 5; Length 390;
Best Local Similarity 86.8%; Pred. NO. 0.023;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttccatggttcagaaccatgtatcattcatcgtggttaaccattc 113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 TCGATGTTGCCATGGTTCTGATCCATGCATCCATCCATCGTGGTAAACCATTC 105

Search completed: April 8, 2000, 02:40:03
Job time: 11118 sec

```

MPRELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:52:11 2000; MasPar time 27.56 Seconds
Tabular output not generated. 15.169 Million cell updates/sec

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.377; Variance 28.586; scale 0.923

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	106	100.0	146	1	DER2_DERPT	MITE ALLERGEN DER P 2	1.34e+14
2	85	80.2	146	1	DEF2_DERFA	MITE ALLERGEN DER F 2	2.19e+08
3	55	51.9	548	1	YDID_ECOLI	HYPOTHETICAL 60.8 KD P	1.13e+00
4	55	51.9	912	1	NIA2_HORVU	NITRATE REDUCTASE (EC	1.13e+00
5	55	51.9	915	1	NIA1_HORVU	NITRATE REDUCTASE (EC	1.13e+00
6	54	50.9	257	1	YJE3_YEAST	HYPOTHETICAL 29.2 KD P	1.89e+00
7	54	50.9	396	1	AMOA_AERHY	PUTATIVE ISOCHORISMATE	1.89e+00
8	54	50.9	397	1	YBP2_YEAST	HYPOTHETICAL 45.5 KD P	1.89e+00
9	54	50.9	787	1	OBP_HSV7J	REPLICATION ORIGIN BIN	1.89e+00
10	54	50.9	1184	1	DP3A_MYCTU	PROBABLE DNA POLYMERASE	1.89e+00
11	53	50.0	278	1	NIFH_METVO	NITROGENASE IRON PROTE	3.12e+00
12	53	50.0	279	1	NIFH_METJA	NITROGENASE IRON PROTE	3.12e+00
13	53	50.0	424	1	OAT_YEAST	ORNITHINE AMINOTRANSFE	3.12e+00
14	53	50.0	631	1	ETED_YEAST	PROBABLE ELECTRON TRAN	3.12e+00
15	52	49.1	188	1	LCAT_PIG	PHOSPHATIDYLCHOLINE-ST	5.10e+00
16	52	49.1	287	1	TISY_MYCPN	THYMIDYLATE SYNTHASE (5.10e+00
17	52	49.1	438	1	LCAT_MOUSE	PHOSPHATIDYLCHOLINE-ST	5.10e+00
18	52	49.1	440	1	LCAT_RABIT	PHOSPHATIDYLCHOLINE-ST	5.10e+00
19	52	49.1	440	1	LCAT_PAPAN	PHOSPHATIDYLCHOLINE-ST	5.10e+00
20	52	49.1	440	1	LCAT_HUMAN	PHOSPHATIDYLCHOLINE-ST	5.10e+00
21	51	48.1	286	1	CBPK_PEA	SERINE CARBOXYPEPTIDAS	8.30e+00
22	51	48.1	309	1	YQGH_BACSU	PROBABLE ABC TRANSPORT	8.30e+00
23	51	48.1	332	1	COBS_PSEDE	COBS PROTEIN.	8.30e+00

24	51	48.1	440	1	LCAT_RAT	PHOSPHATIDYLCHOLINE-ST	8.30e+00
25	51	48.1	655	1	YQGH_ECOLI	HYPOTHETICAL 70.0 KD P	8.30e+00
26	51	48.1	1176	1	CARB_TRICU	CARBAMOYL-PHOSPHATE SY	8.30e+00
27	51	48.1	1607	1	LMG1_MOUSE	LAMININ GAMMA-1 CHAIN	8.30e+00
28	51	48.1	2499	1	MPRI_BOVIN	CATION-INDEPENDENT MAN	8.30e+00
29	50	47.2	325	1	SVK3_ECOLI	HYPOTHETICAL LYSYL-TRN	1.34e+01
30	50	47.2	329	1	GCC_RAT	IG GAMMA-2C CHAIN C RE	1.34e+01
31	50	47.2	355	1	FK38_HUMAN	-38 KD FK-506 BINDING P	1.34e+01
32	50	47.2	370	1	PSPB_RABIT	PULMONARY SURFACTANT-A	1.34e+01
33	50	47.2	373	1	Y007_METJA	HYPOTHETICAL PROTEIN M	1.34e+01
34	50	47.2	387	1	YB96_YEAST	HYPOTHETICAL 43.3 KD P	1.34e+01
35	50	47.2	432	1	Y175_PYRHO	HYPOTHETICAL 43.3 KD P	1.34e+01
36	50	47.2	483	1	TR2_HUMAN	ORPHAN NUCLEAR RECEPT	1.34e+01
37	50	47.2	588	1	RYG2_YEAST	RETROGRADE REGULATION	1.34e+01
38	50	47.2	778	1	TAST_HUMAN	TASTIN (TROPHININ-ASSI	1.34e+01
39	49	46.2	209	1	GTP_MESAU	GLUTATHIONE S-TRANSFER	2.14e+01
40	49	46.2	209	1	GTP1_MOUSE	GLUTATHIONE S-TRANSFER	2.14e+01
41	49	46.2	222	1	R11C_TOBAC	RAS-RELATED PROTEIN RA	2.14e+01
42	49	46.2	284	1	NIF1_METYL	NITROGENASE IRON PROTE	2.14e+01
43	49	46.2	292	1	NIF2_METYL	NITROGENASE IRON PROTE	2.14e+01
44	49	46.2	441	1	YLIG_ECOLI	HYPOTHETICAL 49.6 KD P	2.14e+01
45	49	46.2	1609	1	LMG1_HUMAN	LAMININ GAMMA-1 CHAIN	2.14e+01

ALIGNMENTS

RESULT 1
ID DER2_DERPT STANDARD; PRT; 146 AA.
AC P49278:
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
GN DERP2.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopitiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RX MEDLINE; 90256301.
RA CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
RA THOMAS W.R.;
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
RT plaque immunoassay";
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
RN [2]
RP PARTIAL SEQUENCE OF 18-57.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATTS-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE; 98409423.
RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
RT "Tertiary structure of the major house dust mite allergen Der p 2:
RT sequential and structural homologies";
RL Biochemistry 37:12707-12714(1998).
CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
DB: 1A9V; 14-OCT-98
KW Allergen; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER P 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;
Query Match 100.0%; Score 106; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 1.34e+14;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 HEIKKVLPGCHGS 41
 QY 1 HEIKKVLPGCHGS 14
 |||||

RESULT 2 STANDARD; PRT; 146 AA.
 ID DEF2_DERFA
 AC Q00855; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERF2.
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphidae;
 OC Dermatophagoides.
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 91291341.
 RA YUUKI T., OKUMURA Y., ANDO T., YAMAKAWA H., SUKO M., HAIDA M.,
 RA OKUDAIRA H.;
 RT "Cloning and expression of cDNA coding for the major house dust mite
 RT allergen Der f II in *Escherichia coli*.";
 RL Agric. Biol. Chem. 55:1233-1238(1991).
 RN [1]
 RN SEQUENCE OF 4-146 FROM N.A.
 RX MEDLINE; 94256850.
 RA OKUHARA H.;
 RT "Molecular biology of mite antigens.";
 RL Alerugi 43:435-440(1994).
 RN [3]
 RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE; 93283958.
 RA NISHIYAMA C., YUUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
 RT "Determination of three disulfide bonds in a major house dust mite
 RT allergen, Der f II.";
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).
 RN [4]
 RP PARTIAL SEQUENCE OF 18-52.
 RX MEDLINE; 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATT'S-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE; 98079068.
 RA ICHIKAWA S., HATANAKA H., YUUKI T., IWAMOTO N., KOJIMA S.,
 RA NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
 RT "Solution structure of Der f 2, the major mite allergen for atopic
 RT diseases.";
 RL J. Biol. Chem. 273:356-360(1998).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC ENBL; D10447; BAA01239.1;
 CC ENBL; D10448; BAA01240.1;
 CC ENBL; D10449; BAA01241.1;
 CC ENBL; S70378; AAB30829.1;

DR PIR; PS0417; PS0417,
 DR PDB; 1AHK; 08-APR-98.
 DR PDB; 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 FT VARIANT 93 93
 FT VARIANT 105 105 M -> V (IN CLONE 1).
 FT VARIANT 128 128 I -> A (IN CLONE 11).
 FT VARIANT 142 142 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMW (IN REF. 2).
 SQ SEQUENCE 146 AA; 72623F23 CRC32;
 Query Match 80.2%; Score 85; DB 1; Length 146;
 Best Local Similarity 78.6%; Pred. No. 2.19e-08;
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 28 NEIKKVLPGCHGS 41
 QY 1 HEIKKVLPGCHGS 14
 :|||||:|||||

RESULT 3
 ID YDID_ECOLI STANDARD; PRT; 548 AA.
 AC P38135; P76202; P76902; P76903;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL 60.8 KD PROTEIN IN AROD-PPSA INTERGENIC REGION.
 GN YDID.
 OS Escherichia coli
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA ATBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K.,
 RA KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,
 RA KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,
 RA MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
 RA SAITO N., SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,
 RA YAMAMOTO Y., YANO M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 356-548 FROM N.A.
 RA HOLZSCHU D.L., MCELVER J.A., LIAO C.C., BERRY A.;
 RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP IDENTIFICATION.
 RX MEDLINE; 95075659.
 RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: TO ENZYMS WHICH ACT VIA AN ATP-DEPENDENT COVALENT
 CC BINDING OF AMP TO THEIR SUBSTRATE. STRONG, TO 2,3-
 CC DIHYDROXYBENZOATE-AMP LIGASE.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB00265; AAC74771.1; ALT_INIT.
CC EMBL; D90812; CAB21391.1; ALT_INIT.
CC EMBL; D90813; CAB21407.1; ALT_INIT.
CC EMBL; M69116; ; NOT_ANNOTATED_CDS.
CC DR HSP; P08659; ILCI.
CC DR ECGENE; EGI2357; YDID.
CC DR PROSITE; PS00455; AMP_BINDING; 1.
CC DR PFAM; PF00501; AMP-binding; 1.
CC KW Hypothetical protein.
CC SEQUENCE 548 AA; 60773 MW; CD545939 CRC32;

Query Match          51.9%; Score 55; DB 1; Length 548;
Best Local Similarity 50.0%; Pred. No. 1.13e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 372 DARKTLPGCG 383
QY 2 EIKVLVPGCHG 13

RESULT 4
ID NIA2_HORVU STANDARD; PRT; 912 AA.
AC P27969;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE NITRATE REDUCTASE (EC 1.6.6.1) (NR) (FRAGMENT).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. STEPTOR; TISSUE=LEAF;
RX MEDLINE; 91326031.
RA SCHNORR K.M., JURICEK M., HUANG C., CULLEY D., KLEINHOF A.;
RT "Analysis of barley nitrate reductase cDNA and genomic clones.";
RL Mol. Gen. Genet. 227:411-416(1991).
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC GROUP IS CALLED CYTOCHROME B-557.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X57844; CAA40975.1;
PIR; S17454; RDBHNS.
HSP; P17571; 2CND.
PROSITE; PS00191; CYTOCHROME_B5; 1.
DR PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.

```

```

DR PFAM; PF00173; heme_1; 1.
DR PFAM; PF00174; oxidored_molyb; 1.
DR PFAM; PF00175; oxidored_fad; 1.
DR PFAM; PF00970; Cyt_reductase; 1.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
KW Nitrate assimilation; Multigene family.
FT NON_TER 1
FT METAL 186 186 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 240 240 MOLYBDENUM-PTERIN (POTENTIAL).
FT FT DISULFID 425 425 INTERCHAIN (POTENTIAL).
FT BINDING 570 570 HEME LIGAND (BY SIMILARITY).
FT BINDING 593 593 HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 912 AA; 101467 MW; 0B3FE625 CRC32;

Query Match          51.9%; Score 55; DB 1; Length 912;
Best Local Similarity 60.0%; Pred. No. 1.13e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 291 RVLPGCIGG 300
QY 5 KVLVPGCHGS 14

RESULT 5
ID NIA1_HORVU STANDARD; PRT; 915 AA.
AC P27967;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE NITRATE REDUCTASE (EC 1.6.6.1) (NR).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. HIMALAYA; TISSUE=LEAF;
RX MEDLINE; 91326031.
RA SCHNORR K.M., JURICEK M., HUANG C., CULLEY D., KLEINHOF A.;
RT "Analysis of barley nitrate reductase cDNA and genomic clones.";
RL Mol. Gen. Genet. 227:411-416(1991).
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC GROUP IS CALLED CYTOCHROME B-557.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X57845; CAA40976.1;
PIR; S17453; RDBHNH.
HSP; P17571; 2CND.
PROSITE; PS00191; CYTOCHROME_B5; 1.
DR PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
DR PFAM; PF00173; heme_1; 1.
DR PFAM; PF00174; oxidored_molyb; 1.
DR PFAM; PF00175; oxidored_fad; 1.
DR PFAM; PF00970; Cyt_reductase; 1.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
KW Nitrate assimilation; Multigene family.

```

```

FT METAL 189 189 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 243 243 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 428 428 INTERCHAIN (POTENTIAL).
FT BINDING 573 573 HEME LIGAND (BY SIMILARITY).
FT BINDING 596 596 HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 915 AA; 101770 MW; F9319B2D CRC32;

Query Match 51.9%; Score 55; DB 1; Length 915;
Best Local Similarity 60.0%; Pred. No. 1.13e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 294 RVLPGCIGG 303
QY 5 KVLVPGCHG 14

RESULT 6
ID YJE3_YEAST STANDARD; PRT; 257 AA.
AC P47053;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 29.2 KD PROTEIN IN GYP6-MHP1 INTERGENIC REGION.
GN YJL043W OR J1204.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA FOHL T.M., ALJINOVIC G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YK015C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49318; CAA89334.1; -
CC KW Hypothetical protein.
CC SEQUENCE 257 AA; 29238 MW; A53A3882 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 257;
Best Local Similarity 72.7%; Pred. No. 1.89e+00;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 39 EILKSLVPKCH 49
QY 2 EIKKLVPGCH 12

RESULT 7
ID AMOA_AERYH STANDARD; PRT; 396 AA.
AC P23300;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE ISOCHORISMATE SYNTHASE (EC 5.4.99.6) (AMONABACTIN).
GN AMOA.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC Aeromonas.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-495A2;
RX MEDLINE; 91317731.
RA BARGHOOUTH S., PAYNE S.M., ARCENEAUX J.E., BYERS B.R.;
RT "Cloning, mutagenesis, and nucleotide sequence of a siderophore
RT biosynthetic gene (amoA) from Aeromonas hydrophila.";
RL J. Bacteriol. 173:5121-5128(1991).

```

```

CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.
CC -1- PATHWAY: AMONABACTIN BIOSYNTHESIS. AMONABACTIN IS AN IRON-
CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE
CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TRPE AND PABB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63339; AAA21935.1; -
CC PIR; A40365; A40365
CC PFAM; PF00425; chorismate_bind; 1.
CC KW Isomerase; Iron transport.
CC SEQUENCE 396 AA; 42074 MW; 7E8C32AD CRC32;

Query Match 50.9%; Score 54; DB 1; Length 396;
Best Local Similarity 54.5%; Pred. No. 1.89e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 247 EIRRVLPKCH 257
QY 2 EIKKLVPGCH 12

RESULT 8
ID YBP2_YEAST STANDARD; PRT; 397 AA.
AC P38226;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 45.5 KD PROTEIN IN FIG1-GIP1 INTERGENIC REGION.
GN YBR042C OR YBR0412.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C;
RA ANDRE B., CZIEPLUCH C., HEIN C., JAUNIAUX J.-C., URRESTARAZU A.,
RA VISSERS S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE 1-ACYL-SN-GLYCEROL-3-PHOSPHATE
CC ACYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z35911; CAA84984.1; -
CC PIR; S45900; S45900.
CC PFAM; PF01553; Acyltransferase; 1.
CC KW Hypothetical protein; Phospholipid biosynthesis; Transferase;
CC Acyltransferase; Transmembrane.
CC TRANSMEM 13 33 POTENTIAL.
CC TRANSMEM 58 78 POTENTIAL.
CC TRANSMEM 118 138 POTENTIAL.
CC TRANSMEM 377 397 POTENTIAL.
CC SEQUENCE 397 AA; 45515 MW; CCAD3795 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 397;
Best Local Similarity 60.0%; Pred. No. 1.89e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```



```
Db 8 HKVRKVVVPG 17
QY 1 HEKKVLVPG 10

RESULT 9
ID OBP_HSV7J STANDARD; PRT; 787 AA.
AC P52379;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE REPLICATION ORIGIN BINDING PROTEIN (OBP).
GN U73
OS Herpes simplex virus (type 7 / strain JI) (HSV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RA NICHOLAS J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC HSV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U43400; AAC54734.1; -.
CC DNA replication: DNA-binding; ATP-binding.
CC NP_BIND 52 59
CC SEQUENCE 787 AA; 90925 MW; 594C2D56 CRC32;
CC -----
Query Match 50.9%; Score 54; DB 1; Length 787;
Best Local Similarity 60.0%; Pred. No. 1.89e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 447 FTKMLVQGGH 456
QY 3 IKKVLVPGCH 12

RESULT 10
ID DP3A_MYCTU STANDARD; PRT; 1184 AA.
AC Q10779;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROBABLE DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).
GN DNAE OR DNAEL OR RV1547 OR MTCY48.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROCHER R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. I.II, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SULLSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```


FT NP_BIND 8 15 ATP (POTENTIAL).
 FT METAL 94 94 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 130 130 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 278 AA; 30396 MW; F9866631 CRC32;

Query Match 50.0%; Score 53; DB 1; Length 278;
 Best Local Similarity 87.5%; Pred. No. 3.12e+00;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 30 KKVLVGCG 37
 ||||| ||
 QY 4 KKVLVGCG 11

RESULT 12
 ID NIFH_METJA STANDARD; PRT; 279 AA.
 AC Q58289;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NITROGENASE IRON PROTEIN (EC 1.18.6.1) (NITROGENASE COMPONENT II)
 DE (NITROGENASE REDUCTASE).
 GN NIFH OR MJ0879.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
 RA SCOTT J.L., GEORGEHAGEN N.S.M., WEIDMAN J.E., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273.1058-1073 (1996).

CC -1- FUNCTION: THE KEY ENZYMATIC REACTIONS IN NITROGEN FIXATION ARE
 CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 3 REDUCED FERREDOXIN + 6 H(+) + N(2) + N ATP =
 CC 3 OXIDIZED FERREDOXIN + 2 NH(3) + N ADP + N ORTHOPHOSPHATE.
 CC -1- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NIFH/FRXC FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67532; AAB98883.1; -;
 CC HSP; P0459; IN2C.
 CC TIGR; MJ0879; -;
 CC PROSITE; PS00692; NIFH_FRXC_2; 1.
 CC PROSITE; PS00746; NIFH_FRXC_1; 1.
 CC PFAM; PF00142; fer4_NifH; 1.
 CC Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.
 NP_BIND 8 15
 FT METAL 94 94 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 130 130 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 279 AA; 30167 MW; 00B2854B CRC32;

Query Match 50.0%; Score 53; DB 1; Length 279;
 Best Local Similarity 87.5%; Pred. No. 3.12e+00;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 30 KKVLVGCG 37
 ||||| ||
 QY 4 KKVLVGCG 11

RESULT 13
 ID OAT_YEAST STANDARD; PRT; 424 AA.
 AC P07991;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ORNITHINE AMINOTRANSFERASE (EC 2.6.1.13) (ORNITHINE--OXO-ACID
 DE AMINOTRANSFERASE).
 GN CAR2 OR CARGB OR YLR438W OR L9753.2.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88055042.
 RA DEGOLS G.;
 RT "Functional analysis of the regulatory region adjacent to the carbB
 RT gene of Saccharomyces cerevisiae. Nucleotide sequence, gene fusion
 RT experiments and cis-dominant regulatory mutation analysis.";
 RL Eur. J. Biochem. 169:193-200(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
 RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,
 RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
 RA JOHNSON D., JOHNSTON L., LANGSTON Y., LATREILLE P., LE T.,
 RA MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
 RA RIEKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
 RA WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-55 FROM N.A.
 RC STRAIN-SIGMA 1278B;
 RX MEDLINE; 87246605.

RA DEGOLS G., JAUNIAUX J.-C., WIAME J.M.;
 RT "Molecular characterization of transposable-element-associated
 RT mutations that lead to constitutive L-ornithine aminotransferase
 RT expression in Saccharomyces cerevisiae";
 RL Eur. J. Biochem. 165:289-296(1987).
 CC -1- CATALYTIC ACTIVITY: L-ORNITHINE + A 2-OXO ACID - L-GLUTAMATE
 CC 5-S-SEMIALDEHYDE + AN L-AMINO ACID.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X06790; CAA29947.1; -;
 CC EMBL; U21094; AAB67514.1; -;
 CC EMBL; X05571; CAA29081.1; -;
 CC PIR; S00181; XNBYO.
 CC HSP; P04181; 1OAT.
 CC SGD; L0000217; CAR2.
 CC PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
 CC PFAM; PF00202; aminotran_3; 1.
 CC Transferase; Aminotransferase; Pyridoxal phosphate.
 KW TRANSFERASE; AMINOTRANSFERASE; PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 FT BINDING 272 272 PYRIDOXAL PHOSPHATE (IN REF. 1 AND 3).
 FT CONFLICT 8 8 MISSING (IN REF. 1 AND 3).
 FT CONFLICT 38 38 V -> L (IN REF. 1 AND 3).
 FT

```

FT CONFLICT 99 99 F -> S (IN REF. 1).
FT CONFLICT 212 212 G -> R (IN REF. 1).
FT CONFLICT 385 385 H -> Q (IN REF. 1).
SQ SEQUENCE 424 AA; 46086 MW; 3F432867 CRC32;

Query Match
Best Local Similarity 50.0%; Score 53; DB 1; Length 424;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 288 HDIMSCFTFGSHGS 301
QY 1 HEIKVLVPGCHGS 14

RESULT 14
ID ETFL_YEAST STANDARD; PRT; 631 AA.
AC Q08822;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-UBIQUINONE OXIDOREDUCTASE
DE PRECURSOR (EC 1.5.1.1) (ETF-QO) (ETF-UBIQUINONE OXIDOREDUCTASE) (ETF
DE DEHYDROGENASE) (ELECTRON-TRANSFERRING-FLAVOPROTEIN DEHYDROGENASE).
GN YOR356W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA DELIUS H., HEILING U., HOFMANN B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: REDUCED ETF + UBIQUINONE -> ETF + UBIQUINOL.
CC -!- COFACTOR: FAD AND A 4FE-4S CLUSTER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ETF-QO / FIXC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 275264; CAA99685.1;
KW Oxidoreductase; Electron transport; Flavoproteins; FAD; Iron-sulfur;
KW 4Fe-4S; Mitochondrion; Transist peptide; Ubiquinone.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 631 PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-
FT UBIQUINONE OXIDOREDUCTASE.
FT NP_BIND 65 79 FAD (ADP PART) (POTENTIAL).
FT METAL 574 574 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 600 600 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 603 603 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 606 606 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 631 AA; 69634 MW; 6CE9A3B3 CRC32;

Query Match
Best Local Similarity 50.0%; Score 53; DB 1; Length 631;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 242 VLAEGCHGS 250
QY 6 VLVPGCHGS 14

RESULT 15
ID LCAT_PIG STANDARD; PRT; 188 AA.
AC P30930;
DT 01-JUL-1993 (Rel. 26, Created)

```

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-
DE CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
DE ACYLTRANSFERASE) (FRAGMENTS).
GN LCAT.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE.
RC TISSUE-PLASMA;
RX MEDLINE; 90075705.
RA YUEKSEL K.U., PARK Y.B., JUNG J., GRACY R.W., LACKO A.G.;
RT "Studies on the structure of lecithin:cholesterol acyltransferase
RT (LACT) -- comparisons of the active site region and secondary
RT structure of the human and the porcine enzymes.";
RL Comp. Biochem. Physiol. 94B:389-394(1989).
RN [2]
RP SEQUENCE OF 97-106 AND 142-154.
RC TISSUE-PLASMA;
RX MEDLINE; 87156771.
RA PARK Y.B., YUEKSEL K.U., GRACY R.W., LACKO A.G.;
RT "The catalytic center of lecithin:cholesterol acyltransferase:
RT isolation and sequence of diisopropyl fluorophosphate-labeled
RT peptides.";
RL Biochem. Biophys. Res. Commun. 143:360-363(1987).
CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL -> STEROL ESTER +
CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
CC ACT AS ACCEPTOR).
CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
CC THIS ENZYME.
CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
DR PIR; PLO153; PLO153.
DR PIR; A29544; A29544.
DR PIR; B29544; B29544.
DR PROSITE; PS00120; LIPASE_SER; PARTIAL.
KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein.
FT CARBOHYD 20 20
FT NON_CONS 34 35
FT NON_CONS 44 45
FT NON_CONS 60 61
FT NON_CONS 66 67
FT NON_CONS 77 78
FT NON_CONS 84 85
FT NON_CONS 96 97
FT NON_CONS 106 107
FT NON_CONS 115 116
FT NON_CONS 154 155
FT NON_TER 188 188
SQ SEQUENCE 188 AA; 21232 MW; 876EB48E CRC32;

Query Match
Best Local Similarity 49.1%; Score 52; DB 1; Length 188;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 26 ILVPGCLGN 34
QY 6 VLVPGCHGS 14

Search completed: Sun Apr 2 17:52:43 2000
Job time : 32 secs.

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:51:06 2000; MasPar time 44.19 Seconds
Tabular output not generated. 14.944 Million cell updates/sec

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r2
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 25.741; Variance 32.196; scale 0.800

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	100%	146	2	A60381 allergen Der p II pre	2.37e-12
2	85	80.2	129	2	A61501 allergen Der f II - h	6.62e-07
3	85	80.2	129	2	JU0394 allergen Der f II (pf)	6.62e-07
4	85	80.2	138	2	A61241 allergen Der f II pre	6.62e-07
5	85	80.2	138	2	B61241 allergen Der f II pre	6.62e-07
6	57	53.8	77	2	A69906 hypothetical protein	1.71e+00
7	55	51.9	250	2	F72247 methionine aminopepti	4.32e+00
8	55	51.9	566	2	E64928 probable acid--CoA li	4.32e+00
9	55	51.9	912	1	RDBHNS nitrate reductase (NA	4.32e+00
10	55	51.9	915	1	RDBHNS nitrate reductase (NA	4.32e+00
11	54	50.9	257	2	S56815 hypothetical protein	6.80e+00
12	54	50.9	396	1	A40365 siderophore biosynthe	6.80e+00
13	54	50.9	397	2	S45900 probable membrane pro	6.80e+00
14	54	50.9	482	1	B69054 DNA-directed DNA poly	6.80e+00
15	54	50.9	960	2	JE0356 gamma-aminobutyric ac	6.80e+00
16	54	50.9	1184	2	H70761 probable dnapolimeras	6.80e+00
17	53	50.0	278	1	NIMXVO nitrogenase (EC 1.18.	1.06e+01
18	53	50.0	279	2	G64409 nitrogenase (EC 1.18.	1.06e+01
19	53	50.0	424	1	XNBYO ornithine--oxo-acid t	1.06e+01
20	53	50.0	426	2	T04527 hypothetical protein	1.06e+01
21	53	50.0	442	2	S50062 cell wall glycoprotei	1.06e+01
22	53	50.0	631	2	S67268 hypothetical protein	1.06e+01
23	53	50.0	632	2	S64786 hypothetical protein	1.06e+01

RESULT ENTRY TITLE	24	52	49.1	110	2	B72730	1.65e+01
25	49.1	141	2	B30868	hypothetical protein	1.65e+01	
26	49.1	179	2	C60328	hypothetical protein	1.65e+01	
27	52	49.1	182	E72312	conserved hypothetical	1.65e+01	
28	52	49.1	188	PL0153	phosphatidylcholine--	1.65e+01	
29	52	49.1	328	S73842	thymidylate synthase	1.65e+01	
30	52	49.1	438	XXMSN	phosphatidylcholine--	1.65e+01	
31	52	49.1	440	JCL502	phosphatidylcholine--	1.65e+01	
32	52	49.1	440	1	XXHUN	1.65e+01	
33	52	49.1	1194	2	S70415	DNA-directed RNA poly	
34	51	48.1	48	2	S55783	laminin gamma chain -	
35	51	48.1	141	2	C69018	methyl viologen-reduc	
36	51	48.1	248	2	F72262	ubiquinone/menaquinon	
37	51	48.1	249	2	E71649	hypothetical protein	
38	51	48.1	286	2	S62370	serine-type carboxype	
39	51	48.1	415	2	T07706	hypothetical protein	
40	51	48.1	440	1	XXRTN	phosphatidylcholine--	
41	51	48.1	505	2	T15159	hypothetical protein	
42	51	48.1	654	2	T10521	beta-glucosidase (EC	
43	51	48.1	655	2	C65243	hypothetical 70.1 kD	
44	51	48.1	1607	1	MMWSB2	laminin gamma-1 chain	
45	51	48.1	2499	1	A30788	mannose 6-phosphate r	

ALIGNMENTS

RESULT 1

ENTRY A60381 #type complete
TITLE allergen Der p II precursor - house-dust mite
(Dermatophagoides pteronyssinus)
ORGANISM #formal_name Dermatophagoides pteronyssinus
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 13-Sep-1998

ACCESSIONS A60381
REFERENCE A60381
#authors Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.; Stewart, G.A.; Thomas, W.R.
#journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
#title Isolation of cDNA coding for the major mite allergen Der p II by IgE plaque immunoassay.
#cross-references MUID:90256301

#accession A60381
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-146 #label CHU

CLASSIFICATION #superfamily allergen Der p II

FEATURE 1-17
18-146
SUMMARY #domain signal sequence #status predicted #label SIG\ #product allergen Der p II #status predicted #label MAT #length 146 #molecular-weight 15999 #checksum 25

Query Match 100.0%; Score 106; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 2.37e-12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 HEIKKVLVPGCHGS 41

QY 1 HEIKKVLVPGCHGS 14

RESULT 2

ENTRY A61501 #type fragment
TITLE allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998

ACCESSIONS A61501
REFERENCE A61501
#authors Trudinger, M.; Chua, K.Y.; Thomas, W.R.
#journal Clin. Exp. Allergy (1991) 21:33-37
#title cDNA encoding the major mite allergen Der f II.
#cross-references MUID:91215495

```

#accession A61501
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-129 #label TRU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476

Query Match 80.2% Score 85; DB 2; Length 129;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 11 NEIKKVMVDCGCHS 24
QY 1 HEIKKVLVPGCHS 14

RESULT 3
ENTRY JU0394 #type complete
TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
ACCESSIONS JU0394
REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.
#journal Agric. Biol. Chem. (1991) 55:1233-1238
#title Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.
#cross-references MUID:91291341
#accession JU0394
#molecule_type mRNA
#residues 1-129 #label YUU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 80.2% Score 85; DB 2; Length 129;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 11 NEIKKVMVDCGCHS 24
QY 1 HEIKKVLVPGCHS 14

RESULT 4
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS A61241; PS0417
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession A61241
#molecule_type mRNA
#residues 1-138 #label YUU
#note part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
10-138 #product allergen Der f II #status experimental #label MAT
SUMMARY #length 138 #checksum 2894

```

```

Query Match 80.2% Score 85; DB 2; Length 138;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 NEIKKVMVDCGCHS 33
QY 1 HEIKKVLVPGCHS 14

RESULT 5
ENTRY B61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS B61241; JU0395
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession B61241
#molecule_type mRNA
#residues 1-138 #label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
10-138 #product allergen Der f II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 80.2% Score 85; DB 2; Length 138;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 NEIKKVMVDCGCHS 33
QY 1 HEIKKVLVPGCHS 14

RESULT 6
ENTRY A69906 #type complete
TITLE hypothetical protein yojC - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS A69906
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

```

```

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitznegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#cross-references MUID:98044033
#accession A69906
#status preliminary: nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-77 #label KUN
#cross-references GB:299114; GB:AL009126; NID:92634230; PID:el185421;
PID:92634342
#experimental_source strain 168
GENETICS
#gene yojC
#summary #length 77 #molecular-weight 8518 #checksum 9771
Query Match 53.8%; Score 57; DB 2; Length 77;
Best Local Similarity 66.7%; Pred. No. 1.71e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 14 RTVLVPACH 22
: |||||
QY 4 KRLVPGCH 12
RESULT 7
ENTRY #type complete
TITLE methionine aminopeptidase - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
23-Jul-1999
ACCESSIONS F72247
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession F72247
#status preliminary
#molecule_type DNA
#residues 1-250 #label ARN
#cross-references GB:AE001798; GB:AE000512; NID:94982033; PID:94982042;
TIGR:TM1478
#experimental_source strain MSB8
GENETICS
#gene TM1478
#summary #superfamily Escherichia coli methionyl aminopeptidase
#length 250 #molecular-weight 27489 #checksum 7507
Query Match 51.9%; Score 55; DB 2; Length 250;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 25 REVRKIVPG 34

```

```

: |||||
QY 1 HEIKVLVPG 10
RESULT 8
ENTRY #type complete
TITLE probable acid-CoA ligase (EC 6.2.1.-) ydid - Escherichia
colli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
26-Aug-1999
ACCESSIONS E64928
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession E64928
#status preliminary: nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-566 #label BLAT
#cross-references GB:AE000265; GB:U00096; NID:g2367122; PID:gl787993;
UWGP:b1701
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ydid
#summary #superfamily 4-coumarate--CoA ligase; acetate--CoA ligase
#length 566 #molecular-weight 62759 #checksum 7145
Query Match 51.9%; Score 55; DB 2; Length 566;
Best Local Similarity 50.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 390 DARTLPPGCEG 401
: |||||
QY 2 EIKKVLVPGCHG 13
RESULT 9
ENTRY #type fragment
TITLE nitrate reductase (NADH) (EC 1.6.6.1) - barley (cv. Steptoe)
(fragment)
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
11-Jun-1999
ACCESSIONS S17454
REFERENCE S17453
#authors Schnorr, K.M.; Juricek, M.; Huang, C.; Culley, D.; Kleinhofs,
A.
#journal Mol. Gen. Genet. (1991) 227:411-416
#title Analysis of barley nitrate reductase cDNA and genomic clones.
#cross-references MUID:91326031
#accession S17454
#status nucleic acid sequence not shown; translation not shown
#molecule_type mRNA
#residues 1-912 #label SCH
#cross-references EMBL:X57844; NID:g19044; PIDN:CAAA0975.1; PID:g19045
#note the nucleotide sequence was submitted to the EMBL Data
Library, February 1991
GENETICS
#map_position 6
#summary #superfamily nitrate reductase (NADH); cytochrome b5 core
homology; cytochrome-b5 reductase homology;
molybdopterin-binding domain homology

```

KEYWORDS chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron; metalloprotein; molybdenum; molybdopterin; NAD; nitrate assimilation; oxidoreductase; phosphoprotein

FEATURE 82-474 #domain molybdopterin-binding domain homology #label PCO\

535-609 #domain cytochrome b5 core homology #label CB5\
658-912 #domain cytochrome-b5 reductase homology #label CBR\
186 #binding_site molybdopterin (Cys) (covalent) #status predicted\
425 #disulfide_bonds interchain #status predicted\
570,593 #binding_site heme iron (His) (axial ligands) #status predicted\
722,884 #binding_site NAD (Lys, Cys) #status predicted\
763 #binding_site FAD (Tyr) #status predicted
SUMMARY #length 912 #checksum 1642

Query Match 51.9%; Score 55; DB 1; Length 912;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 291 RVLPGCIGG 300
QY 5 KVLVPGCHGS 14
:||:|||||:

RESULT 10 RDBNH #type complete
ENTRY nitrate reductase (NADH) (EC 1.6.6.1) - barley (cv. Himalaya)
TITLE #formal_name Hordeum vulgare #common_name barley
ORGANISM 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
DATE 11-Jun-1999

ACCESSIONS S17453
REFERENCE S17453
#authors Schnorr, K.M.; Juricek, M.; Huang, C.; Culley, D.; Kleinhofs, A.

#journal Mol. Gen. Genet. (1991) 227:411-416
#title Analysis of barley nitrate reductase cDNA and genomic clones.
#cross-references EMBL:X57845; NID:g18993; PIDN:CAA40976.1; PID:g18994
#note the nucleotide sequence was submitted to the EMBL Data Library, February 1991

GENETICS #map_position 6
#introns 391/1
CLASSIFICATION #superfamily nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 reductase homology;
molybdopterin-binding domain homology

KEYWORDS chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron; metalloprotein; molybdenum; molybdopterin; NAD; nitrate assimilation; oxidoreductase; phosphoprotein

FEATURE 85-477 #domain molybdopterin-binding domain homology #label PCO\
538-612 #domain cytochrome b5 core homology #label CB5\
661-915 #domain cytochrome-b5 reductase homology #label CBR\
189 #binding_site molybdopterin (Cys) (covalent) #status predicted

428 #disulfide_bonds interchain #status predicted\
573,596 #binding_site heme iron (His) (axial ligands) #status predicted\
725,887 #binding_site NAD (Lys, Cys) #status predicted\
766 #binding_site FAD (Tyr) #status predicted
SUMMARY #length 915 #molecular-weight 10170 #checksum 3558

Query Match 51.9%; Score 55; DB 1; Length 915;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 294 RVLPGCIGG 303
QY 5 KVLVPGCHGS 14
:||:|||||:

RESULT 11 S56815 #type complete
ENTRY hypothetical protein YJL043w - yeast (Saccharomyces cerevisiae)
TITLE
ALTERNATE_NAMES hypothetical protein J1204
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 14-Nov-1997

ACCESSIONS S56815
REFERENCE S56793
#authors Pohl, T.M.; Aljinovic, G.
#submission submitted to the Protein Sequence Database, September 1995
#accession S56815
#molecule_type DNA
#residues 1-257 #label TOV
#cross-references EMBL:249318; NID:g1008170; PID:g1008171; MIPS:YJL043w

GENETICS #map_position 10L
SUMMARY #length 257 #molecular-weight 29238 #checksum 4555

Query Match 50.9%; Score 54; DB 2; Length 257;
Best Local Similarity 72.7%; Pred. No. 6.80e+00;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 39 EILKSLVPKCH 49
QY 2 EIKKVLVPGCH 12
|||:|||||

RESULT 12 A40365 #type complete
ENTRY siderophore biosynthetic protein amoA - Aeromonas hydrophila
TITLE #formal_name Aeromonas hydrophila
ORGANISM 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
DATE 10-Sep-1999

ACCESSIONS A40365
REFERENCE A40365
#authors Barghouti, S.; Payne, S.M.; Arceneaux, J.E.L.; Byers, B.R.
#journal J. Bacteriol. (1991) 173:5121-5128
#title Cloning, mutagenesis, and nucleotide sequence of a siderophore biosynthetic gene (amoA) from Aeromonas hydrophila.

#cross-references MUID:91317731
#accession A40365
#status preliminary
#molecule_type DNA
#residues 1-396 #label BAR
#cross-references GB:M63339
#note the authors translated the codon GAG for residue 393 as Gly

CLASSIFICATION #superfamily isochorismate synthase
SUMMARY #length 396 #molecular-weight 42042 #checksum 6097

Query Match 50.9%; Score 54; DB 1; Length 396;
Best Local Similarity 54.5%; Pred. No. 6.80e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 247 EIRRVLTPCR 257
QY 2 EIKKVLVPGCH 12
||:|||||:

RESULT 13 S45900 #type complete
ENTRY probable membrane protein YBR042c - yeast (Saccharomyces cerevisiae)
TITLE
ALTERNATE_NAMES hypothetical protein YBR0412
ORGANISM #formal_name Saccharomyces cerevisiae

```

DATE          26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
ACCESSIONS    S45900
REFERENCE      S45893
#authors      Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.;
#submission   Urrestazu, A.; Vissers, S.
#accession    S45900 submitted to the Protein Sequence Database, August 1994
#molecule_type DNA
#residues     1-397 #label AND
#cross-references EMBL:Z35911; NID:G536265; PID:CAA84984.1;
#experimental_source strain S288C
GENETICS
#map_position 2R
CLASSIFICATION #superfamily probable membrane protein YBR042c
KEYWORDS       transmembrane protein
FEATURE
12-37          #domain transmembrane #status predicted #label TM1\
55-77          #domain transmembrane #status predicted #label TM2\
134-150        #domain transmembrane #status predicted #label TM3\
372-390        #domain transmembrane #status predicted #label TM4\
SUMMARY        #length 397 #molecular-weight 45515 #checksum 7874
Query Match.   50.9%; Score 54; DB 2; Length 397;
Best Local Similarity 60.0%; Pred. No. 6.80e+00;
Matches        6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db            8 HKVRKVVVPG 17
QY            1 HEIKKVLVPG 10

RESULT 14
ENTRY   B69054 #type complete
TITLE   DNA-directed DNA polymerase (EC 2.7.7.7) delta small chain -
ORGANISM Methanobacterium thermoautotrophicum (strain Delta H)
DATE     05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS B69054
REFERENCE   A69000
#authors    Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
            Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
            Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
            Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
            Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
            A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
            McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
            Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
            J.; Reeve, J.N.
#journal    J. Bacteriol. (1997) 179:7135-7155
#title      Complete genome sequence of Methanobacterium
            thermoautotrophicum Delta H: functional analysis and
            comparative genomics.
#cross-references MUID:98037514
#accession    B69054
#status       nucleic acid sequence not shown; translation not shown
#residues     1-482 #label MTH
#cross-references GB:AE000903; GB:AE000666; NID:g2622514;
            PID:AA85882.1; PID:g2622517
#experimental_source strain Delta H
GENETICS
#gene        MTH1405
CLASSIFICATION #superfamily DNA-directed DNA polymerase delta small chain;
            phosphoesterase core homology
KEYWORDS      metalloprotein; nucleotidyltransferase
FEATURE
223-321      #domain phosphoesterase core homology #label PEC
SUMMARY        #length 482 #molecular-weight 54375 #checksum 3013
Query Match   50.9%; Score 54; DB 1; Length 482;

```

```

Best Local Similarity 53.8%; Pred. No. 6.80e+00;
Matches          7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db            310 DIKIVMIPGRHDS 322
QY            2 EIRKVLVPGCHGS 14

RESULT 15
ENTRY   JE0356 #type complete
TITLE   gamma-aminobutyric acid receptor B precursor - human
ALTERNATE_NAMES GABA(B) receptor
ORGANISM #formal_name Homo sapiens #common_name man
DATE      05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change
ACCESSIONS JE0356
REFERENCE   JE0356
#authors    Grifa, A.; Totaro, A.; Rommens, J.M.; Carella, M.; Roetto,
            A.; Borgato, L.; Zelante, L.; Gasparini, P.;
            Biochem. Biophys. Res. Commun. (1998) 250:240-245
            GABA (gamma-amino-butyric acid) neurotransmission:
            Identification and fine mapping of the human GABAB receptor
            gene.
#cross-references MUID:98440782
#accession    JE0356
#molecule_type mRNA
#residues     1-960 #label GRI
#cross-references GB:Y11044; NID:g2826760
#note         this ORF is not annotated in GenBank entry HSGTHLAL,
            release 109
GENETICS
#map_position 6p21.3-6p21.3
KEYWORDS      glycoprotein; neurotransmitter receptor; transmembrane
            protein
FEATURE
1-11         #domain signal sequence #status predicted #label SIG\
12-960       #product gamma-aminobutyric acid receptor B #status
            predicted #label MAT\
590-613      #domain transmembrane #status predicted #label TM1\
627-654      #domain transmembrane #status predicted #label TM2\
666-687      #domain transmembrane #status predicted #label TM3\
709-730      #domain transmembrane #status predicted #label TM4\
767-788      #domain transmembrane #status predicted #label TM5\
803-825      #domain transmembrane #status predicted #label TM6\
831-856      #domain transmembrane #status predicted #label TM7\
23,83,439,481,501, #binding_site carbohydrate (Asn) (covalent) #status
            513,630      Predicted
SUMMARY        #length 960 #molecular-weight 108148 #checksum 3766
Query Match   50.9%; Score 54; DB 2; Length 960;
Best Local Similarity 66.7%; Pred. No. 6.80e+00;
Matches        6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db            237 IKITLMPGC 245
QY            3 IKKVLVPGC 11

Search completed: Sun Apr 2 17:51:55 2000
Job time : 49 secs.

```

MPERCH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:47:58 2000; MacPar time 67.33 Seconds
31.923 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-1
Description: (1-31) from US09362731.pep
Perfect Score: 224

Sequence: 1 QYIKANSFIGITELGGHEIKKVLPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl12

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 31.412; Variance 45.412; scale 0.692

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	101	45.1	145	5	GROUP 2 ALLERGEN EUR M	1.10e-06
2	74	33.0	635	14	K5 LYASE.	2.64e-01
3	73	32.6	205	2	SCO2 PROTEIN PRECURSOR	4.01e-01
4	72	32.1	210	10	FUCOXANTHIN CHLOROPHYL	6.07e-01
5	72	32.1	471	5	SPTPR4 (EC 3.1.3.48) (6.07e-01
6	70	31.3	208	5	T26C5.1 PROTEIN.	1.37e-00
7	70	31.3	393	1	393AA LONG HYPOTHETICA	1.37e-00
8	70	31.3	841	10	PUTATIVE POLYPROTEIN	1.37e-00
9	68	30.4	189	2	PUTATIVE TRANSCRIPTION	3.07e-00
10	68	30.4	224	14	CAPSID PROTEIN (FRAGME	3.07e-00
11	67	29.9	291	2	HYDROXYMETHYLGLICINE	4.55e+00
12	67	29.9	485	3	HEXOKINASE PI (HKX2).	4.55e+00
13	67	29.9	517	10	OSNRMP1.	4.55e+00
14	67	29.9	692	2	ELONGATION FACTOR G (E	4.55e+00
15	66	29.5	118	2	RPON RNA POLYMERASE SI	6.73e-00
16	66	29.5	134	10	RIBOSOMAL PROTEIN S14.	6.73e-00
17	66	29.5	234	2	HYPOTHETICAL 25.9 KD P	6.73e-00
18	66	29.5	308	14	GENOME, PARTIAL SEQUE	6.73e-00
19	66	29.5	320	5	COSMID T28A11.	6.73e-00
20	66	29.5	442	2	BETAINE REDUCTASE.	6.73e+00

21	66	29.5	449	2	Q927T8	REPLICATIVE DNA HELICA	6.73e+00
22	66	29.5	458	5	Q4506	F42G8.9 PROTEIN.	6.73e+00
23	66	29.5	637	2	Q59440	ENDOGLUCANASE 3 PRECUR	6.73e+00
24	66	29.5	1012	5	Q21075	KOLAG.1 PROTEIN.	6.73e+00
25	66	29.5	1408	1	Q27557	CELL SURFACE GLYCOPROT	6.73e+00
26	66	29.5	1463	5	Q4384	PRGAG-POL.	6.73e+00
27	65	29.0	69	2	O86200	MCHI PROTEIN PRECURSOR	9.91e+00
28	65	29.0	131	7	Q48669	MHC CLASS II BETA CHAI	9.91e+00
29	65	29.0	148	1	Q53997	SEFD.	9.91e+00
30	65	29.0	155	1	Q9YDC5	155AA LONG HYPOTHETICA	9.91e+00
31	65	29.0	207	9	Q02582	O PROTEIN.	9.91e+00
32	65	29.0	207	9	Q38271	LECITHIN:CHOLESTEROL A	9.91e+00
33	65	29.0	440	11	O35849	DNA REPAIR PROTEIN (SM	9.91e+00
34	65	29.0	455	2	O83985	PHOSPHORYLASE PHOSPHAT	9.91e+00
35	65	29.0	589	13	Q92138	PUTATIVE TRANSCRIPTION	9.91e+00
36	65	29.0	697	2	Q45419	SENSORY TRANSDUCTION H	9.91e+00
37	65	29.0	1178	2	O55693	RIBOSOMAL PROTEIN S14.	1.45e+01
38	64	28.6	151	5	O9XQ04	CONSERVED HYPOTHETICAL	1.45e+01
39	64	28.6	157	3	O74797	XYLUKINASE.	1.45e+01
40	64	28.6	248	1	O59175	TGTP1.	1.45e+01
41	64	28.6	493	2	O05181	C17G1.3 PROTEIN.	1.45e+01
42	64	28.6	510	5	Q27072	LEGUMIN-LIKE PROTEIN.	1.45e+01
43	64	28.6	530	5	O93242	YEAB PROTEIN.	1.45e+01
44	64	28.6	565	10	O23878		
45	64	28.6	598	2	O34469		

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	145 AA.
ID	O96430			
AC	O96430;			
DT	01-MAY-1999 (TREMBlrel. 10, Created)			
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)			
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)			
DE	GROUP 2 ALLERGEN EUR M 2 0101.			
GN	EUR M 2 0101.			
OS	Euroglyphus maynei (House-dust mite).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;			
OC	Acariiformes; Sarcophagiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;			
OC	Euroglyphus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	SMITH W., HART B.J., THOMAS W.R.;			
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF047613; AAC82349.1;			
DR	HSSP; P49278; 1A9V.			
SQ	SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;			

Query Match 45.1%; Score 101; DB 5; Length 145;
Best Local Similarity 59.1%; Pred. No. 1.10e-06;
Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db	19	VDIKDCANHEIKKVMVPGCKGS 40
Qy	10	IGITELGGHEIKKVLPGCHGS 31

RESULT	2	PRELIMINARY;	PRT;	635 AA.
ID	O09496			
AC	O09496;			
DT	01-JUL-1997 (TREMBlrel. 04, Created)			
DT	01-JUL-1997 (TREMBlrel. 04, Last sequence update)			
DT	01-AUG-1998 (TREMBlrel. 07, Last annotation update)			
DE	K5 LYASE.			
OS	coliphage K5.			
OC	Viruses.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	ESWEH F., HANFLING P., JANN B., JANN K., ROBERTS I.S.;			
RL	Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; Y10025; CAA71133.1;			
KW	Lyase.			

Best Local Similarity 45.5%; Pred. No. 6.07e-01;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 129 FIGLIELGPAIKKELEADCEA 150
|||: ||| :||| | : | :
QY 9 FIGITELGGHEIKKVLPGCHG 30

RESULT 5
ID Q9Y1X6 PRELIMINARY; PRT; 471 AA.
AC Q9Y1X6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE SPTPR4 (EC 3.1.3.48) (FRAGMENT).
OS Ephydatia fluviatilis.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haptosclerida; Spongillidae; Ephydatia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99246376.
RA ONO K., SUGA H., IWABE N., KUMA K., MIYATA T.;
RT "Multiple protein tyrosine phosphatases in sponges and explosive gene
RT duplication in the early evolution of animals before the parazoan-
RT eumetazoan split.";
RL J. Mol. Evol. 48:654-662(1999).
DR EMBL; AB019125; BAA82558.1;
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 471 AA; 54509 MW; BA496490 CRC32;

Query Match 32.1%; Score 72; DB 5; Length 471;
Best Local Similarity 22.6%; Pred. No. 6.07e-01;
Matches 7; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

Db 223 HYNKVNKNTYIVPDSHRCRMSIIPGVGS 253
|||: ||| :||| | : | :
QY 1 QYIKANSKFIGITELGGHEIKKVLPGCHG 31

RESULT 6
ID Q22814 PRELIMINARY; PRT; 208 AA.
AC Q22814;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE T26C5.1 PROTEIN.
GN T26C5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA THOMAS K.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z50859; CAA90726.1; -.

Best Local Similarity 33.0%; Score 74; DB 14; Length 635;
Matches 11; Conservative 13; Mismatches 7; Indels 2; Gaps 2;

Db 273 DYSTSDARQAHGTEGSDNVNVLMSGCDGT 305
:| :||| | :| :||| :||| | :
QY 1 QYIKANSKFI-GITELGGHE-IRKVLPGCHGS 31

RESULT 3
ID Q92CW7 PRELIMINARY; PRT; 205 AA.
AC Q92CW7;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE SCO2 PROTEIN PRECURSOR (SCO2).
GN RP587.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE; 99039499.
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAEGLUND A.K.,
RA ERIKSSON A.S., WINKLER H.H., KUHLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RA ANDERSSON S.G.E.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ235272; CAA15032.1; -.

Query Match 32.68%; Score 73; DB 2; Length 205;
Best Local Similarity 45.8%; Pred. No. 4.01e-01;
Matches 11; Conservative 6; Mismatches 5; Indels 2; Gaps 2;

Db 126 EYIKNFHPKFTSLTG-NEHQIKDV 148
|||: ||| :||| | : | :
QY 1 QYIKANSKFIGITELGGHEIKKVLPGCHG 31

RESULT 4
ID Q39969 PRELIMINARY; PRT; 210 AA.
AC Q39969;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE FUCOXANTHIN CHLOROPHYLL A / C BINDING PROTEIN.
GN FCPL.
OS Heterosigma carterae.
OC Eukaryota; stramenopiles; Raphidophyceae; Heterosigma.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97156944.
RA DURNFORD D.G., AEBERSOLD R., GREEN B.R.;
RT "The fucoxanthin-chlorophyll proteins from a chromophyte alga are part
RT of a large multigene family: structural and evolutionary relationships
RT to other light harvesting antennae.";
RL Mol. Gen. Genet. 253:377-386(1996).
RL EMBL; X99697; CAA68028.1; -.
DR MENDEL; 8468; Hetica; 1394:8468.
DR PFAM; PF00504; Chloro_a_b_bind; 1.
SQ SEQUENCE 210 AA; 23238 MW; 892C5808 CRC32;

Query Match 32.1%; Score 72; DB 10; Length 210;

SQ SEQUENCE 635 AA; 66944 MW; 58219280 CRC32;

Query Match 33.0%; Score 74; DB 14; Length 635;
Best Local Similarity 33.3%; Pred. No. 2.64e-01;
Matches 11; Conservative 13; Mismatches 7; Indels 2; Gaps 2;

Db 273 DYSTSDARQAAGVTEGSDNINNVLMSGCDGT 305
| : |||| | : |||| | : |||| | :
QY 1 QYIKANSKFI-GITELGGHE-IKKVLPGCHGS 31

RESULT 3
ID Q92CW7 PRELIMINARY; PRT; 205 AA.
AC Q92CW7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE SPTPR4 (EC 3.1.3.48) (FRAGMENT).
OS Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
GN RP587.
NC Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE; 99039499.
RA ANDERSSON S.G.E.; ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA ERIKSSON A.S., WINKLER H.H., KUHLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RA ANDERSSON S.G.E.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ235272; CRA15032.1;
SQ SEQUENCE 205 AA; 23644 MW; 6A4E199C CRC32;

Query Match 32.6%; Score 73; DB 2; Length 205;
Best Local Similarity 45.8%; Pred. No. 4.01e-01;
Matches 11; Conservative 6; Mismatches 5; Indels 2; Gaps 2;

Db 126 EYKNFHPKFISLTG-NEHQIKDV 148
| : |||| | : |||| | : |||| | :
QY 1 QYIKA-NSKFIGITELGGHEIKKV 23

RESULT 4
ID Q39969 PRELIMINARY; PRT; 210 AA.
AC Q39969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE FUCOXANTHIN CHLOROPHYLL A / C BINDING PROTEIN.
GN FCPL.
OS Heterostigma carterae.
OC Eukaryota; stramenopiles; Raphidophyceae; Heterosigma.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97156944.
RA DURNFORD D.G.; AEBERSOLD R.; GREEN B.R.;
RT "The fucoxanthin-chlorophyll proteins from a chromophyte alga are part
of a large multigene family: structural and evolutionary relationships
to other light harvesting antennae";
RL Mol. Gen. Genet. 253:377-386(1996).
RL EMBL; X99697; CAA68028.1;
DR MENDEL; 8468; Hetica; 1394:8468.
DR PFAM; PF00504; Chloro_a_b_wmd; 1.
SQ SEQUENCE 210 AA; 23238 MW; 892C5808 CRC32;

Query Match 32.1%; Score 72; DB 10; Length 210;

SQ SEQUENCE 841 AA; 96367 MW; E7C2968A CRC32;
 Query Match 31.3%; Score 70; DB 10; Length 841;
 Best Local Similarity 42.1%; Pred. No. 1.37e+00;
 Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 264 TKVFAELGDNVEKALV 282
 :|:::||||:|:
 QY 7 SKFIGITELGGHEIKKVLV 25

RESULT 9
 ID Q9ZBM6 PRELIMINARY; PRT; 189 AA.
 AC Q9ZBM6;
 DT 01-WAY-1999 (TREMBLrel. 10, Created)
 DT 01-WAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-WAY-1999 (TREMBLrel. 10, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.
 DE MLCB1450.06C.
 OS Mycobacterium leprae.
 GN Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BROWN D., CHURCHER C.M.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA JAMES K.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93188700.
 RA EIGLMETER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
 RT "Use of an ordered cosmid library to deduce the genomic organization
 RT of Mycobacterium leprae."
 RT Mol. Microbiol. 7:197-206(1993).
 DR EMBL; AL035159; CAA22690.1; -
 SQ SEQUENCE 189 AA; 19894 MW; 42073499 CRC32;

Query Match 30.4%; Score 68; DB 2; Length 189;
 Best Local Similarity 44.4%; Pred. No. 3.07e+00;
 Matches 8; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Db 149 TDGVREVKSILV-GCOA 165
 :||:|:||:|:
 QY 13 TELGGHEIKKLVPGCHG 30

RESULT 10
 ID P89671 PRELIMINARY; PRT; 224 AA.
 AC P89671;
 DT 01-WAY-1997 (TREMBLrel. 03, Created)
 DT 01-WAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
 DE CAPSID PROTEIN (FRAGMENT).
 OS San Miguel sea lion virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
 OC Calicivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NEILL J.D., MEYER R.F., SEAL B.S.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U76886; AAB48432.1; -
 FT NON_TER 1
 FT NON_TER 224 224
 SQ SEQUENCE 224 AA; 24382 MW; 75AA14BA CRC32;

Query Match 30.4%; Score 68; DB 14; Length 224;
 Best Local Similarity 38.9%; Pred. No. 3.07e+00;
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 115 KSQSVYTADLGNTVK 132

QY 4 KANSFIGITELGGHEIK 21

```

Query Match      29.9%; Score 67; DB 2; Length 291;
Best Local Similarity 38.7%; Pred. No. 4.55e+00;
Matches 12; Conservative 9; Mismatches 9; Indels 1; Gaps 1;

Db      209 DYFKALDSKEVRTTVEAESFMKVLNGGCHS 239
ov      :|||:||||| : : :||| :|||
         1 OYIEA-NSKFGITFELGGHRTKVLVPGCHG 10

```

```

Query Match      29.9% ; Score 67 ; DB 3 ; Length 486 ;
Best Local Similarity 40.9% ; Pred. No. 4.55e+00 ;
Matches      9 ; Conservative      8 ; Mismatches 4 ; Indels      1 ; Gaps      1 ;

```

```

RESULT      14
ID          Q9ZK24      PRELIMINARY;      PRT;      692 AA.
AC          Q9ZK24;
DT          01-MAY-1999 (TReMBLrel. 10, Created)
DT          01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT          01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE          ELONGATION FACTOR G (EF-G).
GN          FUSA.
OS          Helicobacter pylori J99.
OC          Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC          Helicobacter.
[1]
RN          SEQUENCE FROM N.A.
RP          STRAIN=J99;
RX          MEDLINE; 99120557.
RA          ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA          SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA          TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA          GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA          TRUST T.J.;
RT          "Genomic-sequence comparison of two unrelated isolates of the human
RT          gastric pathogen Helicobacter pylori.";
RL          Nature 397:176-180(1999).
DR          ENBL; AE001539; AAD06689.1; -.
DR          HSSP; P13551; 2FFG.
DR          PROSITE; PS00301; EFACITOR_GTP; 1.
KW          Protein biosynthesis; GTP-binding.
SQ          SEQUENCE      692 AA;  77127 MW;  05C6CBEC CRC32;

```

```

Query Match      29.9%  Score 67;  DB 2;  Length 692;
Best Local Similarity 47.6%  Pred. No. 4.55e+00;
Matches 10;  Conservative 3;  Mismatches 8;  Indels 0;  Gaps 0;

Db      229  KYLGGEELDIYEETKKGIKTGC 249
Ov      8  KFIGITELGGGEITKKVLVPGC 28

```

Search completed: Sun Apr 2 17:49:11 2000
Job time : 73 secs.

RELEASE

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:47:09 2000; Maspar time 28.18 Seconds
Tabular output not generated. 32.853 Million cell updates/sec

Title: >US-09-362-731-1
Description: (1-31) from US09362731.pep
Perfect Score: 224
Sequence: 1 QYKANSKFIGITELGGHEIKKVLVPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 32.909; Variance 46.340; scale 0.710

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	113	50.4	146	1	DER2_DERPT MITE ALLERGEN DER P 2	1.79e-09
2	104	46.4	1314	1	TETX_CLOTE TETANUS TOXIN PRECURSO	1.72e-07
3	92	41.1	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	5.89e-05
4	74	33.0	417	1	DHML_PARDE METHYLAMINE DEHYDROGEN	1.82e-01
5	71	31.7	382	1	TGFI_XENLA TRANSFORMING GROWTH FA	6.26e-01
6	69	30.8	287	1	TRUE_AQUAE TRNA PSEUDOURIDINE SYN	1.40e+00
7	69	30.8	364	1	DP3B_MYCPN DNA POLYMERASE III, BE	1.40e+00
8	69	30.8	380	1	DP3B_MYCPN DNA POLYMERASE III, BE	1.40e+00
9	69	30.8	618	1	NOP2_YEAST NUCLEOLAR PROTEIN NOP2	1.40e+00
10	69	30.8	631	1	EFG_HELPY ELONGATION FACTOR G (E	1.40e+00
11	68	30.4	1882	1	Y468_MYCPN HYPOTHETICAL PROTEIN M	2.07e+00
12	67	29.9	310	1	PYRB_SALTY ASPARTATE CARBAMOYLTRA	3.06e+00
13	67	29.9	485	1	HXBK_YEAST HEXOKINASE B (EC 2.7.1	3.06e+00
14	66	29.5	140	1	YB17_YEAST HYPOTHETICAL 15.8 KD P	4.50e+00
15	66	29.5	188	1	LCAT_PIG PHOSPHATIDYLCHOLINE-ST	4.50e+00
16	66	29.5	213	1	KAD_MYCCA ADENYLATE KINASE (EC 2	4.50e+00
17	66	29.5	438	1	LCAT_MOUSE PHOSPHATIDYLCHOLINE-ST	4.50e+00
18	66	29.5	440	1	LCAT_PAPAN PHOSPHATIDYLCHOLINE-ST	4.50e+00
19	66	29.5	440	1	LCAT_HUMAN PHOSPHATIDYLCHOLINE-ST	4.50e+00
20	66	29.5	440	1	LCAT_RABIT PHOSPHATIDYLCHOLINE-ST	4.50e+00
21	66	29.5	808	1	SYFE_SYN7 PHENYLALANYL-TRNA SYN	4.50e+00
22	66	29.5	1035	1	POLY_DROME RETROVIRUS-RELATED POL	4.50e+00
23	65	29.0	129	1	EXBD_HAEDU BIOPOLYMER TRANSPORT E	6.58e+00

24	65	29.0	145	1	EXBD_PASHA BIOPOLYMER TRANSPORT E	6.58e+00
25	65	29.0	207	1	REGQ_LAMBD ANTITERMINATION PROTEIN	6.58e+00
26	65	29.0	221	1	Y805_METUA HYPOTHETICAL PROTEIN M	6.58e+00
27	65	29.0	226	1	TRPF_METUA N-(5'-PHOSPHORIBOSYL)A	6.58e+00
28	65	29.0	305	1	PYRB_SERMA ASPARTATE CARBAMOYLTRA	6.58e+00
29	65	29.0	339	1	GPDA_ECOLI GLYCEROL-3-PHOSPHATE D	6.58e+00
30	65	29.0	379	1	HYPD_ALCEU HYDROGENASE EXPRESSION	6.58e+00
31	65	29.0	440	1	LCAT_RAT PHOSPHATIDYLCHOLINE-ST	6.58e+00
32	65	29.0	455	1	RADA_TREPA DNA REPAIR PROTEIN RAD	6.58e+00
33	65	29.0	485	1	HXKA_YEAST HEXOKINASE A (EC 2.7.1	6.58e+00
34	65	29.0	487	1	HUS2_YEAST DNA REPAIR PROTEIN	6.58e+00
35	65	29.0	614	1	DNAK_ODOSI DNAK PROTEIN (HEAT SHO	6.58e+00
36	64	28.6	151	1	RS14_DROME 40S RIBOSOMAL PROTEIN	9.59e+00
37	64	28.6	171	1	GPDA_SALTY GLYCEROL-3-PHOSPHATE D	9.59e+00
38	64	28.6	329	1	NADA_CYPAA QUINOLINATE SYNTHETASE	9.59e+00
39	64	28.6	396	1	YCGA_BACSU HYPOTHETICAL 41.5 KD P	9.59e+00
40	64	28.6	415	1	AMST_ERWAM AMYLOVORAN BIOSYNTHESI	9.59e+00
41	64	28.6	425	1	DHMH_PARVE METHYLAMINE DEHYDROGEN	9.59e+00
42	64	28.6	451	1	MURD_BACSU UDP-N-ACETYLURAMIDYLAL	9.59e+00
43	64	28.6	590	1	2AAA_DROME PROTEIN PHOSPHATASE PP	9.59e+00
44	64	28.6	631	1	XRC1_MOUSE DNA-REPAIR PROTEIN XRC	9.59e+00
45	64	28.6	633	1	XRC1_HUMAN DNA-REPAIR PROTEIN XRC	9.59e+00

ALIGNMENTS

RESULT 1
ID DER2_DERPT STANDARD; PRT; 146 AA.
AC P49278;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
GN DERP2.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90256301.
RA CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
RA THOMAS W.R.;
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
RT plaque immunoassay."
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
RN [2]
RP PARTIAL SEQUENCE OF 18-57.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATT'S-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE; 98409423.
RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
RT "Tertiary structure of the major house dust mite allergen Der p 2:
RT sequential and structural homologies";
RL Biochemistry 37:12707-12714(1998).
CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
DR PDB; 1A9V; 14-OCT-98.
KW Allergen; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER P 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;

Query Match 50.4%; Score 113; DB 1; Length 146;
Best Local Similarity 63.6%; Pred. No. 1.79e-09;

Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 20 VDVKCANHEIKKVLPGCHGS 41
 QY 10 IGITELGGHEIKKVLPGCHGS 31

RESULT 2

ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: primary structure, expression in E. coli, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911.
 RX MEDLINE; 87040747.
 RA FAIRWEATHER N.F., LYNESS V.A.;
 RL "The complete nucleotide sequence of tetanus toxin.";
 RN Nucleic Acids Res. 14:7809-7812(1986).
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE; 86085672.
 RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in Escherichia coli.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE; 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulphydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE; 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]
 RP IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL Nature 359:832-835(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.

RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL Nat. Struct. Biol. 4:788-792(1997).
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
 CC GANGLIOSIDE RECEPTORS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04436; CAA28033.1; -;
 DR EMBL; M12739; AAA3282.1; -;
 DR EMBL; X06214; CAA29564.1; -;
 DR PIR; A25689; BTCLN;
 DR PDB; 1AF9; 29-APR-98.
 DR PDB; 1ABD; 14-OCT-98.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
 KW 3D-structure.
 FT INIT_MET 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 233 233 BY SIMILARITY.
 FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT DISULFID 438 466 INTERCHAIN.
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
 Query Match 46.48; Score 104; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 1.72e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 829 QYIKANSKFIGITEL 843
 QY 1 QYIKANSKFIGITEL 15
 RESULT 3
 ID DEF2_DERFA STANDARD; PRT; 146 AA.
 AC Q00855; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERF2
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
 OC Dermatophagoides.

Query Match 41.1%; Score 92; DB 1; Length 146;
Best Local Similarity 50.0%; Pred. No. 5.89e-05;
Matches 11; Conservative 6; Mismatches 5; Indels

FT HELIX 46 58
 FT TURN 59 60
 FT TURN 75 76
 FT STRAND 77 82
 FT HELIX 84 86
 FT STRAND 90 96
 FT TURN 97 100
 FT STRAND 101 108
 FT STRAND 113 116
 FT TURN 118 119
 FT STRAND 123 132
 FT TURN 133 134
 FT STRAND 135 145
 FT TURN 147 149
 FT STRAND 152 158
 FT TURN 159 160
 FT STRAND 166 166
 FT HELIX 170 172
 FT STRAND 173 175
 FT TURN 177 178
 FT STRAND 181 186
 FT STRAND 192 197
 FT TURN 198 201
 FT STRAND 202 208
 FT STRAND 212 219
 FT TURN 220 221
 FT STRAND 222 227
 FT TURN 228 229
 FT STRAND 232 236
 FT STRAND 244 247
 FT TURN 254 255
 FT STRAND 258 258
 FT STRAND 263 265
 FT TURN 266 269
 FT STRAND 270 275
 FT TURN 276 277
 FT STRAND 279 284
 FT TURN 286 287
 FT STRAND 291 292
 FT STRAND 296 297
 FT HELIX 301 305
 FT TURN 306 307
 FT STRAND 308 310
 FT STRAND 316 319
 FT TURN 320 323
 FT STRAND 324 331
 FT TURN 334 335
 FT TURN 337 338
 FT STRAND 341 348
 FT TURN 349 351
 FT STRAND 354 364
 FT STRAND 366 369
 FT STRAND 376 381
 FT TURN 382 385
 FT STRAND 386 391
 FT TURN 392 394
 FT STRAND 397 401
 FT STRAND 410 412
 SQ SEQUENCE 417 AA; 45440 MW; 89367A39 CRC32;

Query Match 33.0%; Score 74; DB 1; Length 417;
 Best Local Similarity 40.0%; Pred. No. 1.82e-01;
 Matches 8; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 193 VGVVDLEGKAFKRLMDVDC 212
 :|:|:|:|:|:|:|:|:|
 QY 10 IGITELGGHEIKVLPVGC 28

RESULT 5
 ID TGF1_XENLA STANDARD; PRT; 382 AA.
 AC P16176;
 DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)
 DE TRANSFORMING GROWTH FACTOR BETA 1 PRECURSOR (TGF-BETA 1) (TGF-BETA 5).
 OS Xenopus laevis. (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90110090.
 RA KONDATIAH P., SANDS M.J., SMITH J.M., FIELDS A., ROBERTS A.B.,
 RA SPORN M.B., MELTON D.A.;
 RT "Identification of a novel transforming growth factor-beta (TGF-beta
 5) mRNA in Xenopus laevis.";
 RL J. Biol. Chem. 265:1089-1093(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA VEMPATI U.D., KONDATIAH P.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IMPORTANT ROLE IN CERTAIN ASPECTS OF DIFFERENTIATION.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J05180; AAA49968.1;
 DR EMBL; AF009335; AAB64441.1;
 DR EMBL; AF009331; AAB64441.1; JOINED.
 DR EMBL; AF009332; AAB64441.1; JOINED.
 DR EMBL; AF009333; AAB64441.1; JOINED.
 DR EMBL; AF009334; AAB64441.1; JOINED.
 DR PIR; A34929; A34929.
 DR HSP; P01137; 1KLC.
 DR PROSITE; PS00250; TGF-BETA; 1.
 DR PFAM; PF00019; TGF-beta; 1.
 DR PFAM; PF00688; TGF-beta; 1.
 DR Growth factor; Mitogen; Glycoprotein; Signal.
 KW SIGNAL
 FT SIGNAL 1 21
 FT PROPEP 22 270
 FT CHAIN 271 382
 FT DISULFID 277 286
 FT DISULFID 285 348
 FT DISULFID 314 379
 FT DISULFID 318 381
 FT DISULFID 347 347
 FT CARBOHYD 73 73
 FT CARBOHYD 123 123
 FT CARBOHYD 166 166
 FT SITE 234 236
 SQ SEQUENCE 382 AA; 44200 MW; ECDC7789 CRC32;

Query Match 31.7%; Score 71; DB 1; Length 382;
 Best Local Similarity 50.0%; Pred. No. 6.26e-01;
 Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 100 YAKQVYRFESITELEDHEFK 119
 :|:|:|:|:|:|:|:|:|
 QY 2 YIKANSRFIGITELGGHEIK 21

RESULT 6
 ID TRUBAQUAE STANDARD; PRT; 287 AA.
 AC O56922;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE TRNA PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70) (TRNA PSEUDOURIDINE 55

DE SYNTHASE) (PSI55 SYNTHASE) (PSEUDOURIDYLATE SYNTHASE) (URACIL
DE HYDROLYASE).
GN TRUB OR AQ_705.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RC MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus".
RL Nature 392:353-358(1998).
CC -1- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE
CC 5'-PHOSPHATE + H(2)O.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000703; AAC06885.1; -
DR PFAM; PF01509; TRUB_N.1.
KW Lyase; tRNA processing.
SQ SEQUENCE 287 AA; 32259 MW; 633369D0 CRC32;

Query Match 30.8%; Score 69; DB 1; Length 287;
Best Local Similarity 59.18; Pred. No. 1.40e+00;
Matches 13; Conservative 3; Mismatches 4; Indels 2; Gaps 2;

Db 266 DSKFIGIGELKGGVLSPKRLLV 287
:||||| || || ||:|
QY 6 NSKFIGITEL-GGH-EIKKVLV 25

RESULT 7
ID DP3B_MYCGE STANDARD; PRT; 364 AA.
AC P47247;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7).
GN DNAN OR MG001.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium".
RL Science 270:397-403(1995).
RN [2]
RP REVISIONS.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,

RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 267-364 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 94364962.
RA BAILEY C.C., BOTT K.F.;
RT "An unusual gene containing a dnaJ N-terminal box flanks the putative
RT origin of replication of Mycoplasma genitalium".
RL J. Bacteriol. 176:5814-5819(1994).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION AND ATP-
CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-
CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND
CC THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE
CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-
CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,
CC GAMMA, AND DELTA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39679; AAC71217.1; -
DR EMBL; U09251; FAA57069.1; -
DR TIGR; MG001; -
DR PFAM; PF00712; DNA_pol3_beta; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 364 AA; 42399 MW; 0CE3F6A2 CRC32;

Query Match 30.8%; Score 69; DB 1; Length 364;
Best Local Similarity 40.0%; Pred. No. 1.40e+00;
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 339 FQGNKYFLITSKSEPELKLVLPS 363
:||||| || || ||:|
QY 3 IKANSKFIGITELGGHEIKKVLVPG 27

RESULT 8
ID DP3B_MYCPN STANDARD; PRT; 380 AA.
AC Q50313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7).
GN DNAN.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 96177562.
RA HILBERT H., HIMMELREICH R., FLAGENS H., HERRMANN R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
RT Mycoplasma pneumoniae comprising the dnaA region, the atp operon and
RT a cluster of ribosomal protein genes".
RL Nucleic Acids Res. 24:628-639(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 97105885.

RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
 RA HERRMANN R.;
 RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*
 RL *pneumoniae*.";
 CC Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT
 CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-
 CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N PROPHOSPHATE + DNA(N).
 CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND
 CC THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE
 CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-
 CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,
 CC GAMMA, AND DELTA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U34816; AAC43645.1; -
 CC EMBL: AB000017; AB95801.1; -
 CC DR PFAM: PF00712; DNA_pol3_beta.1.
 CC KW Transferase; DNA-directed DNA polymerase; DNA replication.
 CC SQ SEQUENCE 380 AA; 43856 MW; 8464DBD6 CRC32;
 CC
 CC Query Match 30.8%; Score 69; DB 1; Length 380;
 CC Best Local Similarity 30.8%; Pred. No. 1.40e+00;
 CC Matches 8; Conservative 10; Mismatches 8; Indels 0; Gaps 0;
 CC
 CC Db 354 YFQSGNRYFLISSNNEPELKVPS 379
 CC | : : : : : | : : : : :
 CC QY 2 YKANSFEGITELGGHEIKKVLPG 27
 CC
 CC RESULT 9
 CC ID N0P2.YEAST STANDARD; PRT; 618 AA.
 CC AC P40991;
 CC DT 01-FEB-1995 (Rel. 31, Created)
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
 CC DT 15-DEC-1999 (Rel. 39, Last annotation update)
 CC DE NUCLEOLAR PROTEIN N0P2.
 CC GN N0P2 OR YNA1 OR YNL061W OR N2428 OR YNL2428W.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC OC Saccharomycetaceae; Saccharomycetes.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-BJ2168.
 CC RX MEDLINE; 95105226.
 CC RA DE BEUS E., BROCKENBROUGH J.S., HONG B., ARIS J.P.;
 CC RT "Yeast N0P2 encodes an essential nucleolar protein with homology to a
 CC RT human proliferation marker.";
 CC RL J. Cell Biol. 127:1799-1813(1994).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-S288C / FY1676;
 CC RX MEDLINE; 96021608.
 CC RA BERGEZ P., DOIGNON F., CROUZET M.;
 CC RT "The sequence of a 44 420 bp fragment located on the left arm of
 CC RT chromosome XIV from *Saccharomyces cerevisiae*.";
 CC RL Yeast 11:967-974(1995).
 CC RN [3]
 CC RP ERRATUM.
 CC RX MEDLINE; 97060022.

RA BERGEZ P., DOIGNON F., CROUZET M.;
 RL Yeast 12:297-297(1996).
 CC [4]
 CC SEQUENCE OF 146-618 FROM N.A.
 CC STRAIN-S288C;
 CC RA GARCIA-BARRIO M.T., CUESTA R., HINNEBUSCH A.G., TAMAME GONZALEZ M.;
 CC Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD BE INVOLVED IN NUCLEOLAR FUNCTION DURING THE ONSET
 CC OF GROWTH, AND IN THE MAINTENANCE OF NUCLEOLAR STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
 CC -1- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
 CC NOLL/NOP2 (EUKARYOTES) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X82656; CAA57979.1; -
 CC EMBL: U12141; AAA95650.1; -
 CC DR EMBL: Z71337; CAA95934.1; -
 CC DR EMBL: X83512; CAA58502.1; -
 CC DR SGD: L0002535; YNA1.
 CC DR PROSITE: PS01153; NOLL_NOP2_SUN; 1.
 CC DR PFAM: PF01189; Noll_Nop2_Sun; 1.
 CC KW Nuclear protein; Ribosome biogenesis.
 CC FT CONFLICT 577 577 I -> M (IN REF. 2).
 CC SQ SEQUENCE 618 AA; 69812 MW; 0EA01260 CRC32;
 CC
 CC Query Match 30.8%; Score 69; DB 1; Length 618;
 CC Best Local Similarity 36.0%; Pred. No. 1.40e+00;
 CC Matches 9; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
 CC
 CC Db 326 HYIQAASSFLPVLDPHENERIL 350
 CC | : : : : : | : : : : :
 CC QY 1 QYI-KANSKFGITELGGHEIKKVL 24
 CC
 CC RESULT 10
 CC ID EFG_HELPY STANDARD; PRT; 691 AA.
 CC AC P56002;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
 CC DE ELONGATION FACTOR G (EF-G).
 CC GN FUSA OR HP1195.
 CC OS Helicobacter pylori (Campylobacter pylori).
 CC OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC OC Helicobacter.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-26695 / ATCC 700392;
 CC RX MEDLINE; 97394467.
 CC RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.;
 CC RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 CC RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 CC RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
 CC RA McKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 CC RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 CC RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
 CC RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 CC RA VENTER J.C.;
 CC RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 CC RT *pylori*.";
 CC RL Nature 388:539-547(1997).
 CC CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 CC RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

[illegible]

OY 1 QYIKANSKFIGITEL-GGHEIKKVLVP 26

RESULT 13
ID HXKB YEAST STANDARD; PRT; 485 AA.
AC P04807;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HEXOKINASE B (EC 2.7.1.1) (HEXOKINASE PII).
GN HXK2 OR HKB OR HEX1 OR YGL253W OR NRB486.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86120382.
RA STACHELEK C., STACHELEK J., SWAN J., BOTSTEIN D., KONIGSBERG W.;
RT "Identification, cloning and sequence determination of the genes
RL specifying hexokinase A and B from yeast.";
RL Nucleic Acids Res. 14:945-963(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86056943.
RA FROEHLICH K.-U., ENTIAN K.-D., MECKE D.;
RT "The primary structure of the yeast hexokinase PII gene (HXK2) which
RL is responsible for glucose repression.";
RL Gene 36:105-111(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE: 97127827.
RA COISSAC E., MAILLIER E., ROBINEAU S., NETTER P.;
RT "Sequence of a 39,411 bp DNA fragment covering the left end of
RL chromosome VII of Saccharomyces cerevisiae.";
RL yeast 12:1555-1562(1996).
RN [4]
RP SEQUENCE OF 1-246 FROM N.A.
RC STRAIN-W303;
RX MEDLINE: 93311123.
RA BREITWIESER W., PRICE C., SCHUSTER T.;
RT "Identification of a gene encoding a novel zinc finger protein in
RL Saccharomyces cerevisiae.";
RL yeast 9:551-556(1993).
RN [5]
RP SEQUENCE OF 118-126; 175-184 AND 303-313.
RC STRAIN-ATCC 38531 / Y41;
RX MEDLINE: 95255188.
RA NORBECK J., BLOMBERG A.;
RT "Gene linkage of two-dimensional polyacrylamide gel electrophoresis
RL resolved proteins from isogene families in Saccharomyces cerevisiae
RL by microsequencing of in-gel trypsin generated peptides.";
RL Electrophoresis 16:149-156(1995).
RN [6]
RP PHOSPHORYLATION OF SER-14.
RX MEDLINE: 94114477.
RA KRIEDEL T.M., RUSH J., VOJTEK A.B., CLIFTON D., FRAENKEL D.G.;
RT "In vivo phosphorylation site of hexokinase 2 in Saccharomyces
RL cerevisiae.";
RL Biochemistry 33:148-152(1994).
RN [7]
RP PHOSPHORYLATION OF SER-157.
RX MEDLINE: 97199316.
RA HEIDRICH K., OTTO A., BEHLKE J., RUSH J., WENZEL K.W., KRIEDEL T.;
RT "Autophosphorylation-inactivation site of hexokinase 2 in
RL Saccharomyces cerevisiae.";
RL Biochemistry 36:1960-1964(1997).
RN [8]
RP SEQUENCE OF 1-18, AND PHOSPHORYLATION OF SER-14.
RX MEDLINE: 98384167.
RA BEHLKE J., HEIDRICH K., NAUMANN M., MULLER E.-C., OTTO A., REUTER R.,
RA KRIEDEL T.;
RT "Hexokinase 2 from Saccharomyces cerevisiae: regulation of oligomeric

RT structure by in vivo phosphorylation at serine-14.";
RL Biochemistry 37:11989-11995(1998).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE: 78244654.
RA ANDERSON C.M., STENKAMP R.E., STEITZ T.A.;
RT "Sequencing a protein by X-ray crystallography. II. Refinement of
RL yeast hexokinase B co-ordinates and sequence at 2.1-A resolution.";
RL J. Mol. Biol. 123:15-33(1978).
CC -1- FUNCTION: MAIN GLUCOSE PHOSPHORYLATING ENZYME. MAY PLAY A
CC REGULATORY ROLE IN BOTH INDUCTION AND REPRESSION OF GENE
CC EXPRESSION BY GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ATP + D-HEXOSE = ADP + D-HEXOSE 6-PHOSPHATE.
CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
CC INHIBITION BY ATP.
CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/H/HK.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X03483; CAA27203.1; -
DR EMBL; M11181; AAA34697.1; -
DR EMBL; X94357; CAA64134.1; -
DR EMBL; 272775; CAA96973.1; -
DR EMBL; X67787; CAA48003.1; -
DR PIR; B23523; KIBYHB.
DR PIR; S28555; S28555.
DR PDB; 2YHX; 15-JUL-92.
DR SWISS-2DPAGE; P04807; YEAST.
DR YEPD; 8536; -
DR YEPD; 8548; -
DR SGD; L00000834; HXK2.
DR PROSITE; PS00378; HEXOKINASES; 1.
DR PFAM; PF00349; hexokinase; 1.
KW Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW 3D-structure; Phosphorylation.
FT INIT_MET 0
FT BINDING 110 110 ATP (BY SIMILARITY).
FT DOMAIN 151 177 GLUCOSE-BINDING (POTENTIAL).
FT MOD_RES 14 14 PHOSPHORYLATION.
FT MOD_RES 157 157 PHOSPHORYLATION.
FT CONFLICT 28 28 N -> I (IN REF. 1).
FT CONFLICT 32 32 I -> N (IN REF. 2).
FT CONFLICT 60 60 G -> V (IN REF. 1).
FT CONFLICT 196 196 T -> S (IN REF. 1).
FT CONFLICT 420 421 YN -> ST (IN REF. 2).
FT CONFLICT 443 444 TS -> PH (IN REF. 2).
FT CONFLICT 452 452 I -> V (IN REF. 2).
FT CONFLICT 461 461 A -> P (IN REF. 2).
FT HELIX 23 33
FT HELIX 37 55
FT STRAND 65 66
FT STRAND 79 86
FT STRAND 90 99
FT TURN 100 101
FT STRAND 102 110
FT STRAND 112 112
FT TURN 115 118
FT TURN 123 123
FT HELIX 124 141
FT TURN 142 142
FT STRAND 150 155

FT STRAND 164 164
 FT TURN 165 166
 FT STRAND 169 169
 FT TURN 175 176
 FT STRAND 187 187
 FT HELIX 188 189
 FT STRAND 202 208
 FT STRAND 210 221
 FT HELIX 220 221
 FT TURN 223 224
 FT STRAND 225 231
 FT STRAND 235 241
 FT HELIX 243 245
 FT TURN 251 252
 FT STRAND 263 266
 FT TURN 270 275
 FT HELIX 283 291
 FT HELIX 299 304
 FT HELIX 306 322
 FT TURN 323 324
 FT TURN 338 339
 FT TURN 343 343
 FT HELIX 344 351
 FT HELIX 358 368
 FT TURN 369 369
 FT HELIX 374 395
 FT TURN 396 396
 FT HELIX 397 406
 FT TURN 407 407
 FT STRAND 411 416
 FT TURN 418 421
 FT TURN 424 425
 FT HELIX 426 438
 FT STRAND 445 447
 FT TURN 450 454
 FT TURN 458 460
 FT HELIX 461 475
 SEQUENCE 485 AA; 53811 MW; 6C3EE887 CRC32;

Query Match 29.9%; Score 67; DB 1; Length 485;
 Best Local Similarity 40.9%; Pred. No. 3.06e+00;
 Matches 9; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 76 KESGDFLAI-DLGGTNRVLV 96
 QY 4 KANSKFIGITELGGHEIKKVLV 25

RESULT 14
 ID YB17 YEAST STANDARD; PRT; 140 AA.
 AC P38291;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 15.8 KD PROTEIN IN TYR1-SSE2 INTERGENIC REGION.
 GN YBR167C OR YBR1219.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA ENTIAN K.-D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,
 RA NIEGEMANN E., SCHENK-GROENINGER R., SERVOS J., WEHNER E.,
 RA WOLTER R., BRENDEN M., BAUER J., BRAUN H., DERN K., DUESTERHUS S.,
 RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREMS B., PROFT M.,
 RA SIEGERS K., BAUR A., BOLES E., MIOSGA T.,
 RA SCHAAFF-GERTSCHLAGER I., ZIMMERMANN F.K.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z36036; CAA85128.1; --
 DR PIR; S46038; S46038.
 KW Hypothetical protein.
 SQ SEQUENCE 140 AA; 15814 MW; 6A571113 CRC32;

Query Match 29.5%; Score 66; DB 1; Length 140;
 Best Local Similarity 30.8%; Pred. No. 4.50e+00;
 Matches 8; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Db 57 KQSSYVAVLGMG-KAVEKTIALGCH 81
 QY 4 KANSKFIGITELGGHEIKKVLVPGCH 29

RESULT 15
 ID LCAT.PIG STANDARD; PRT; 188 AA.
 AC P30930;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-
 DE CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
 DE ACYLTRANSFERASE) (FRAGMENTS).
 GN LCAT.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-PLASMA;
 RX MEDLINE; 90075705.
 RA YUEKSEL K.U., PARK Y.B., JUNG J., GRACY R.W., LACKO A.G.;
 RT "Studies on the structure of lecithin:cholesterol acyltransferase
 RT (LACAT) -- comparisons of the active site region and secondary
 RT structure of the human and the porcine enzymes.";
 RL Comp. Biochem. Physiol. 94B:389-394(1989).
 RN [2]
 RP SEQUENCE OF 97-106 AND 142-154.
 RC TISSUE-PLASMA;
 RX MEDLINE; 87156771.
 RA PARK Y.B., YUEKSEL K.U., GRACY R.W., LACKO A.G.;
 RT "The catalytic center of lecithin:cholesterol acyltransferase:
 RT Isolation and sequence of diisopropyl fluorophosphate-labeled
 RT peptides.";
 RL Biochem. Biophys. Res. Commun. 143:360-363(1987).
 CC -1- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
 CC ACT AS ACCEPTOR).
 CC -1- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
 CC THIS ENZYME.
 CC -1- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC PIR; P0153; P0153.
 DR PIR; A29544; A29544.
 DR PIR; B29544; B29544.
 DR PROSITE; PS00120; LIPASE_SER; PARTIAL.
 KW Transferase: Acyltransferase; Lipid metabolism; Glycoprotein.
 FT CARBOHYD 20 20
 FT NON_CONS 34 35
 FT NON_CONS 44 45
 FT NON_CONS 60 61
 FT NON_CONS 66 67
 FT NON_CONS 77 78
 FT NON_CONS 84 85
 FT NON_CONS 96 97

FT NON_CONS 106 107
FT NON_CONS 115 116
FT NON_CONS 154 135
FT NON_TER 188 188
SQ SEQUENCE 188 AA; 21232 MW; 876EB48E CRC32;

Query Match 29.5%; Score 66; DB 1; Length 188;
Best Local Similarity 47.4%; Pred. No. 4.50e+00;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 16 AELSNHTRPVILVPGGIGN 34
QY 13 TELGGHEIKKVLVPGCHGS 31

Search completed: Sun Apr 2 17:47:40 2000
Job time : 31 secs.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Sun Apr 2 17:46:02 2000; MasPar time 45.22 Seconds

Tabular output not generated. 32.336 Million cell updates/sec

Title: >US-09-362-731-1
Description: (1-31) from US09362731.ppe
Perfect Score: 224
Sequence: 1 QYKANSKFIGITELGGHEIKKVLPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62

Statistics: 1:pir1 2:pir2 3:pir3 4:pir4

Mean 32.179; Variance 52.202; scale 0.616

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	113	50.4	146	2 A60381	allergen Der p II pre	7.50e-08
2	104	46.4	1315	1 BTCLTN	tentoxylisin (EC 3.4.4.16e-06	4.16e-06
3	92	41.1	129	2 A61501	allergen Der f II - h	7.10e-04
4	92	41.1	129	2 JU0394	allergen Der f II (pF	7.10e-04
5	92	41.1	138	2 A61241	allergen Der f II pre	7.10e-04
6	92	41.1	138	2 B61241	amine dehydrogenase (8.72e-01
7	74	33.0	417	2 JH0560	sco2 protein precursor	1.26e+00
8	73	32.6	205	2 F70361	transforming growth f	2.61e+00
9	71	31.7	382	2 B61036	hypothetical protein	3.74e+00
10	70	31.3	393	2 A71201	DNA polymerase III be	5.33e+00
11	69	30.8	287	2 A64200	tRNA-pseudouridine sy	5.33e+00
12	69	30.8	287	2 F70361	DNA-directed DNA poly	5.33e+00
13	69	30.8	380	2 S62836	nucleolar protein NOP	5.33e+00
14	69	30.8	618	2 A55188	translation elongatio	5.33e+00
15	69	30.8	692	2 C64669	hypothetical protein	7.58e+00
16	68	30.4	1882	2 S73484	co-induced hydrogense	1.07e+01
17	67	29.9	138	2 H51114	aspartate carbamoylitr	1.07e+01
18	67	29.9	311	1 OWE8AC	hexokinase (EC 2.7.1.	1.07e+01
19	67	29.9	486	1 KIBYHA	translating factor g (1.07e+01
20	67	29.9	517	2 S62667	elongation factor 9 (1.07e+01
21	67	29.9	692	2 G71847	hypothetical protein	1.51e+01
22	66	29.5	140	2 S46038	phosphatidylcholine--	1.51e+01
23	66	29.5	188	2 PL0153		

24	66	29.5	213	1 KIYMC	adenylate kinase (EC	1.51e+01
25	66	29.5	234	2 S76694	hypothetical protein	1.51e+01
26	66	29.5	438	1 XXMSN	phosphatidylcholine--	1.51e+01
27	66	29.5	440	1 JC1502	phosphatidylcholine--	1.51e+01
28	66	29.5	440	1 XXHUN	phosphatidylcholine--	1.51e+01
29	66	29.5	449	2 D72056	replicative DNA helic	1.51e+01
30	66	29.5	1035	1 GNFG1	retrovirus-related po	1.51e+01
31	66	29.5	1408	2 H69088	cell surface glycopro	1.51e+01
32	65	29.0	157	2 D72696	hypothetical protein	2.13e+01
33	65	29.0	207	1 ZQBPL	regulatory protein Q	2.13e+01
34	65	29.0	221	2 E64400	conserved hypothetica	2.13e+01
35	65	29.0	226	2 C64356	phosphoribosylanthran	2.13e+01
36	65	29.0	306	1 OWE8AC	aspartate carbamoylitr	2.13e+01
37	65	29.0	339	2 S47829	glycerol-3-phosphate	2.13e+01
38	65	29.0	379	2 S29978	hypD protein - Alkali	2.13e+01
39	65	29.0	440	1 XXRTN	phosphatidylcholine--	2.13e+01
40	65	29.0	455	1 A71251	DNA repair protein sm	2.13e+01
41	65	29.0	485	1 KIBYHA	hexokinase (EC 2.7.1.	2.13e+01
42	65	29.0	487	1 S55194	DNA-directed DNA poly	2.13e+01
43	65	29.0	589	2 S65953	[phosphorylase] phosph	2.13e+01
44	65	29.0	614	2 S78277	dnak-type molecular c	2.13e+01
45	65	29.0	1178	2 S76370	sensory transduction	2.13e+01

ALIGNMENTS

RESULT 1	A60381	#type complete
ENTRY	allergen Der p II precursor - house-dust mite	
TITLE	(Dermatophagoides pteronyssinus)	
ORGANISM	#formal_name Dermatophagoides pteronyssinus	
DATE	03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 13-Sep-1998	
ACCESSIONS	A60381	
REFERENCE	A60381	
#authors	Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;	
#journal	Stewart, G.A.; Thomas, W.R.	
#title	Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123	
#cross-references	Isolation of cDNA coding for the major mite allergen Der p II by IgE plaque immunoassay.	
#accession	#domain signal sequence #status predicted #label SIG\	
#status	#product allergen Der p II #status predicted #label MAT	
#molecule_type	#length 146 #molecular-weight 15999 #checksum 25	
#residues	1-146 #label CHU	
CLASSIFICATION	#superfamily allergen Der p II	
FEATURE	1-17	
1-17	18-146	
SUMMARY	Query Match 50.4%; Score 113; DB 2: Length 146; Best Local Similarity 63.6%; Pred. No. 7.50e-08; Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;	

RESULT	2
ENTRY	BTCLTN #type complete
TITLE	tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani
ALTERNATE_NAMES	tetanus neurotoxin
ORGANISM	#formal_name Clostridium tetani
DATE	31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 18-Jun-1999
ACCESSIONS	A25689; A25757; A25194; B25194; A60759; S69348; S09364
REFERENCE	A25689
#authors	Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weiler, U.; Hudel, M.; Habermann, E.; Niemann, H.
#journal	EMBO J. (1986) 5:2495-2502
#title	Tetanus toxin: primary structure, expression in E. coli, and

homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
#molecule_type DNA
#residues 1-1315 #label EIS
#cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
#molecule_type DNA
#residues 1-1315 #label FAI
#cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
#experimental_source strain CN3911
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
#journal J. Bacteriol. (1986) 165:21-27
#title Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.
#cross-references MUID:86085672
#accession A25194
#molecule_type DNA
#residues 743-1315 #label FA2
#cross-references GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
#accession B25194
#molecule_type protein
#residues 865-894 #label FA3
REFERENCE A60759
#authors Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
#journal Infect. Immun. (1989) 57:3588-3593
#title Isolation, purification, and characterization of fragment B, the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 #label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell epitopes.
#cross-references MUID:89093918
#contents annotation; epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
#molecule_type protein
#residues 2-31 #label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B,C) chains that are covalently linked by an interchain disulfide bond (the individual chains are not toxic when separated). The amino end of the heavy chain (fragment B) can be separated from the carboxyl end (fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds

to gangliosides and may target the toxin to the motor end plate. Fragment A is a zinc-dependent endopeptidase. This potent neurotoxin binds to peripheral neuronal synapses, is internalized, and moves by retrograde transport up the axon into the spinal cord where it can move between postsynaptic and presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin (vesicle-associated membrane protein 2).

FUNCTION #description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
FEATURE 2-457 : #product tentoxylisin light chain (fragment A) #status predicted #label TRU
461-1315 : #product tentoxylisin heavy chain (fragment B,C) #status predicted #label TRU
461-864 : #domain channel forming (fragment B) #status predicted #label TRU
865-1315 : #domain ganglioside binding (fragment C) #status predicted #label TRU
233,237 : #binding_site zinc (His) #status predicted
234 : #active_site Glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 46.4% Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 4.16e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 QYIKANSKFGITEL 844
|||||
QY 1 QYIKANSKFGITEL 15

RESULT 3
ENTRY #type fragment
TITLE allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
ACCESSIONS A61501
REFERENCE #authors Trudinger, M.; Chua, K.Y.; Thomas, W.R.
#journal Clin. Exp. Allergy (1991) 21:33-37
#title CDNA encoding the major mite allergen Der f II.
#cross-references MUID:91215495
#accession A61501
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-129 #label TRU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476
Query Match 41.1% Score 92; DB 2; Length 129;
Best Local Similarity 50.0%; Pred. No. 7.10e-04;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 3 VDVKDCANNEIKKVMVDCGHS 24
: : : : :
QY 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 4
ENTRY #type complete
TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
ACCESSIONS JU0394

PS0417	Yuuiki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.
#journal	Agric. Biol. Chem. (1991) 55:1233-1238
#title	Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.
#cross-references	MUID:91291341
#accession	JU0394
##molecule_type	mRNA
##residues	1-129 ##label YUU
CLASSIFICATION	#superfamily allergen Der p II
SUMMARY	#length 129 #molecular-weight 14076 #checksum 9516
Query Match	41.1%; Score 92; DB 2; Length 129;
Best Local Similarity	50.0%; Pred. No. 7.10e-04;
Matches	11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Db	3 VDVKDCANNEIKKVMVDCGCHS 24 : : : : :
Qy	10 IGITELGGHEIKKVLVPDCHGS 31
RESULT	5
ENTRY	A61241
TITLE	#type fragment allergen Der f ii precursor - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM	#formal_name Dermatophagoides farinae
DATE	12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS	A61241; PS0417
REFERENCE	A61241
#authors	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H. Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#journal	
#title	Synthesis of biologically active recombinant Der f II.
#cross-references	MUID:92040281
#accession	A61241
##molecule_type	mRNA
##residues	1-138 ##label YUU
##note	part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing
CLASSIFICATION	#superfamily allergen Der p II
FEATURE	1-9
10-138	#domain signal sequence (fragment) #status predicted #label sig\
SUMMARY	#product allergen Der f II #status experimental #label MAT #length 138 #checksum 2894
Query Match	41.1%; Score 92; DB 2; Length 138;
Best Local Similarity	50.0%; Pred. No. 7.10e-04;
Matches	11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Db	12 VDVKDCANNEIKKVMVDCGCHS 33 : : : : :
Qy	10 IGITELGGHEIKKVLVPDCHGS 31
RESULT	6
ENTRY	B61241
TITLE	#type fragment allergen Der f ii precursor - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM	#formal_name Dermatophagoides farinae
DATE	12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS	B61241; JU0395
REFERENCE	A61241
#authors	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H. Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#journal	
#title	Synthesis of biologically active recombinant Der f II.
#cross-references	MUID:92040281
#accession	B61241

##molecule_type	mrna
##residues	1-138 ##label YUU
CLASSIFICATION	#superfamily allergen Der p II
FEATURE	1-9
10-138	#domain signal sequence (fragment) #status predicted #label sig\ #product allergen Der f II #status predicted #label MAT #length 138 #checksum 2280
SUMMARY	
Query Match	41.1%; Score 92; DB 2; Length 138;
Best Local Similarity	50.0%; Pred. No. 7.10e-04;
Matches	11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Db	12 VDVKDCANNEIKKVMVDCGCHS 33 : : : : :
Qy	10 IGITELGGHEIKKVLVPDCHGS 31
RESULT	7
ENTRY	JH0660
TITLE	#type complete amine dehydrogenase (EC 1.4.99.3) large chain precursor - Paracoccus denitrificans
ALTERNATE_NAMES	methylamine dehydrogenase large chain
ORGANISM	#formal_name Paracoccus denitrificans
DATE	17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 23-Jul-1999
ACCESSIONS	JH0660; S51049
REFERENCE	PH0856
#authors	Chistoserdov, A.Y.; Boyd, J.; Mathews, F.S.; Lidstrom, M.E.
#journal	Biochem. Biophys. Res. Commun. (1992) 184:1181-1189
#title	The genetic organization of the mau gene cluster of the facultative autotroph Paracoccus denitrificans.
#cross-references	MUID:92272706
#accession	JH0660
##molecule_type	DNA
##residues	1-417 ##label CHI
##cross-references	GB:M90099; NID:g150583; PID:g150585
REFERENCE	S51046
#authors	van Spanning, R.J.M.; van der Palen, C.J.N.M.; Slotboom, D.J.; Reijnders, W.N.M.; Stouthamer, A.H.; Duine, J.A.
#journal	Eur. J. Biochem. (1994) 226:201-210
#title	Expression of the mau genes involved in methylamine metabolism in Paracoccus denitrificans is under control of a LysR-type transcriptional activator.
#cross-references	MUID:95045590
#accession	S51049
##status	preliminary
##molecule_type	DNA
##residues	1-69 ##label VAN
GENETICS	##cross-references EMBL:U12464; NID:g558801; PID:g558805
#gene	maub
FUNCTION	#description catalyzes the oxidation of methylamine to formaldehyde and ammonia oxidoreductase
KEYWORDS	
FEATURE	1-28
29-417	#domain (or 1-26) signal sequence #status predicted #product (or 27-417) amine dehydrogenase large chain #status predicted #label AMI
SUMMARY	#length 417 #molecular-weight 45440 #checksum 4630
Query Match	33.0%; Score 74; DB 2; Length 417;
Best Local Similarity	40.0%; Pred. No. 8.72e-01;
Matches	8; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db	193 VGVVDLEGRFAFKRMULDVPDC 212 : : : : : : : :
Qy	10 IGITELGGHEIKKVL-VPGC 28
RESULT	8

```

ENTRY      F71663      #type complete
TITLE      sco2 protein precursor (sco2) RP587 - Rickettsia prowazekii
ORGANISM   #formal_name Rickettsia prowazekii
DATE       21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
ACCESSIONS F71663
REFERENCE   A71630
#authors   Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
            Sacheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
            Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
            C.G.
#journal   Nature (1998) 396:133-140
#title     The genome sequence of Rickettsia prowazekii and the origin
            of mitochondria.
#cross-references MUID:99039499
#accession F71663
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-205 #label AND
#cross-references GB:AJ235272; GB:AJ235269; NID:g3861033; PID:el342876;
            PID:g3861132
#experimental_source strain Madrid E
GENETICS
#gene      sco2; RP587
CLASSIFICATION
#superfamily immunodominant protein
SUMMARY    #length 205 #molecular-weight 23644 #checksum 748
Query Match 32.6%; Score 73; DB 2; Length 205;
Best Local Similarity 45.8%; Pred. No. 1.26e+00;
Matches 11; Conservative 6; Mismatches 5; Indels 2; Gaps 2;
Db 126 EYKFNHPKFLSLTG-NEHQIKDV 148
      :||| :|||:| :|||
QY 1 QYIKA-NSKFIGITELGGHEIKV 23
      :||| :|||:| :|||

RESULT 9
ENTRY   B61036      #type complete
TITLE   transforming growth factor beta-5 precursor - African clawed
            frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE     31-Dec-1993 #sequence_revision 03-Feb-1994 #text_change
ACCESSIONS A34929; B61036
REFERENCE   A34929
#authors   Kondaiah, P.; Sands, M.J.; Smith, J.M.; Fields, A.; Roberts,
            A.B.; Sporn, M.B.; Melton, D.A.
#journal   J. Biol. Chem. (1990) 265:1089-1093
#title     Identification of a novel transforming growth factor-beta
            (TGF-beta5) mRNA in Xenopus laevis.
#cross-references MUID:90110090
#accession A34929
#molecule_type mRNA
#residues  1-382 #label KON
#cross-references GB:J05180; NID:g214821; PIDN:AAA49968.1; PID:g214822
REFERENCE   A61036
#authors   Roberts, A.B.; Rosa, F.; Roche, N.S.; Coligan, J.E.;
            Garfield, M.; Rebert, M.L.; Kondaiah, P.; Danielpour, D.;
            Kehrl, J.H.; Wahl, S.M.; Dawid, I.B.; Sporn, M.B.
#journal   Growth Factors (1990) 2:135-147
#title     Isolation and characterization of TGF-beta2 and TGF-beta5
            from medium conditioned by Xenopus XTC cells.
#cross-references MUID:90253806
#accession B61036
#molecule_type protein
#residues  271-276,'x',278-284,'xx',287-299 #label ROB
CLASSIFICATION
#superfamily inhibin
#growth factor
#product transforming growth factor beta-5 #status
#experimental #label MAT
SUMMARY    #length 382 #molecular-weight 44200 #checksum 3471

```

```

Query Match 31.7%; Score 71; DB 2; Length 382;
Best Local Similarity 50.0%; Pred. No. 2.61e+00;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 100 YAKQVYRPESITELEDHEFK 119
      :||| :|||:| :|||
QY 2 YIKANSKFIGITELGGHEIK 21
      :||| :|||:| :|||

```

RESULT 10

```

ENTRY   A71201      #type complete
TITLE   hypothetical protein PH1878 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE     14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSIONS A71201
REFERENCE   A71000
#authors   Kwarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
            Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
            Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
            Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
            Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
            A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
            Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal   DNA Res. (1998) 5:55-76
#title     Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
            OT3.
#cross-references MUID:98344137
#accession A71201
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-393 #label KAW
#cross-references GB:AP000007; NID:g3236134; PID:d1031943; PID:g3258317
#experimental_source strain OT3
#note      this accession replaces an interim accession for a
            sequence replaced by GenBank

```

GENETICS

#gene

SUMMARY

```

Query Match 31.3%; Score 70; DB 2; Length 393;
Best Local Similarity 36.4%; Pred. No. 3.74e+00;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 350 QFLIRNAKFLDAIEAENAIK 371
      :||| :|||:| :|||
QY 1 QYIKANSKFIGITELGGHEIKK 22
      :||| :|||:| :|||

```

RESULT 11

```

ENTRY   A64200      #type complete
TITLE   DNA polymerase III beta chain - Mycoplasma genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE     17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
ACCESSIONS A64200
REFERENCE   A64200
#authors   Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
            R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
            Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
            Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
            Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
            J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
            Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
            C.A.; Venter, J.C.
#journal   Science (1995) 270:397-403
#title     The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession A64200
#status    preliminary; nucleic acid sequence not shown;

```

```

#molecule_type DNA
#residues 1-267 #label TIGR
#cross-references GB:U39679; GB:L43967; NID:g1045668; PID:g1045669;
#experimental_source strain G-37
TIGR:MG001

GENETICS
#gene dnaN
#genetic_code SGC3
#start_codon GTG
SUMMARY
#length 267 #molecular_weight 31240 #checksum 4856
Query Match 30.8%; Score 69; DB 2; Length 267;
Best Local Similarity 40.0%; Pred. No. 5.33e+00;
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 242 FQGNKYFLITKSEPELKOILVPS 266
QY 3 IRANSFGITELGGHEIKKVLVPG 27

RESULT 12
ENTRY F70361 #type complete
TITLE tRNA-pseudouridine synthase (EC 5.4.99.-) truD - Aquifex
ALTERNATE_NAMES tRNA pseudouridine 55 synthase
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
24-Sep-1999
ACCESSIONS F70361
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession F70361
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-287 #label AQF
#cross-references GB:AE000703; NID:g2983287; PIDN:AACC6885.1;
#experimental_source strain VF5

GENETICS
#gene truD
CLASSIFICATION #superfamily Escherichia coli protein P35
KEYWORDS intramolecular transferase; isomerase; tRNA modification
SUMMARY #length 287 #molecular_weight 32259 #checksum 7335
Query Match 30.8%; Score 69; DB 2; Length 287;
Best Local Similarity 59.1%; Pred. No. 5.33e+00;
Matches 13; Conservative 3; Mismatches 4; Indels 2; Gaps 2;

Db 266 DSKFIGIGELKGVLSPKRLIV 287
QY 6 NSKFIGITEL-GGH-EIKKVLV 25

RESULT 13
ENTRY S62836 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain dnaN
ALTERNATE_NAMES - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
ORGANISM hypothetical protein K05_orf380
#formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change
18-Sep-1998
ACCESSIONS S62836; S73479
REFERENCE S62797

#authors Hilbert, H.; Himmelreich, R.; Plagens, H.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:628-639
#title Sequence analysis of 56 kb from the genome of the bacterium
Mycoplasma pneumoniae comprising the dnaA region, the atp
operon and a cluster of ribosomal protein genes.
#cross-references MUID:96177562
#accession S62836
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-380 #label HTL
#cross-references EMBL:U34816; NID:g1209514; PID:g1209517
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession S73479
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-380 #label HIM
#cross-references EMBL:AE000017; GB:U00089; NID:g1673812; PID:g1673814
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996
GENETICS
#gene dnaN
#genetic_code SGC3
KEYWORDS nucleotidyltransferase
SUMMARY #length 380 #molecular_weight 43856 #checksum 8925
Query Match 30.8%; Score 69; DB 2; Length 380;
Best Local Similarity 30.8%; Pred. No. 5.33e+00;
Matches 8; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 354 YFQSNKYFLISSNNEPELKEILVPS 379
QY 2 YIKANSFGITELGGHEIKKVLVPG 27

RESULT 14
ENTRY A55188 #type complete
TITLE nucleolar protein NOP2 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein N2428; protein YNL061w; YNAL protein
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS A55188; B55188; S51670; S58717; S62989
REFERENCE A55188
#authors de Beus, E.; Brockenbrough, J.S.; Hong, B.; Aris, J.P.
#journal J. Cell Biol. (1994) 127:1799-1813
#title Yeast NOP2 encodes an essential nucleolar protein with
homology to a human proliferation marker.
#cross-references MUID:95105226
#accession A55188
#molecule_type DNA
#residues 1-618 #label DEA
#cross-references GB:X82656; NID:g576441; PID:g576442
#accession B55188
#molecule_type protein
#residues 207-217:355-369 #label DE2
REFERENCE S51669
#authors Garcia-Barrio, M.T.; Cuesta, R.; Hinnebusch, A.G.; Tamame
Gonzalez, M.
#submission submitted to the EMBL Data Library, December 1994
#accession S51670
#molecule_type DNA
#residues 146-576, 'M', 578-618 #label GAR
#cross-references EMBL:X83512; NID:g603586; PID:g603587

```

```

11-138      #domain translation elongation factor Tu homology #label
            ETVU\
17-24      #region nucleotide-binding motif A (P-loop)\
135-138    #region GTP-binding NKXD motif
SUMMARY    #length 692 #molecular-weight 77021 #checksum 508

Query Match      30.8%; Score 69; DB 2; Length 692;
Best Local Similarity 47.6%; Pred. No. 5.33e+00;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 229 KYLGGEELSIIEIKKGKAGC 249
    |::|::|::|::|::|::|::|
Qy 8 KFIGITELGGHEIKKVLVPGC 28

Search completed: Sun Apr 2 17:46:51 2000
Job time : 49 secs.

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
Run on: Sun Apr 2 17:53:01 2000; MasPar time 66.11 seconds
14.684 Million cell updates/sec
```

```
>US-09-362-731-2
Description: (1-14) from US09362731.pep
perfect Score: 106
Sequence: 1 HEIKKVLVPGCHGS 14
```

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: sptrembl12

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 25.207; Variance 28.345; scale 0.889

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	92	86.8	145	5	O96430	GROUP 2 ALLERGEN EUR M	4.49e+10
2	63	59.4	157	3	O74797	CONSERVED HYPOTHETICAL	2.53e-02
3	61	57.5	282	10	O92900	T8011.21 PROTEIN.	7.66e-02
4	57	53.8	77	2	O31860	YOUC PROTEIN.	6.56e-01
5	57	53.8	697	2	O45419	PUTATIVE TRANSCRIPTION	6.56e-01
6	56	52.8	520	11	O61235	SYNTROPHIN-2.	1.11e+00
7	56	52.8	540	4	Q13425	BETA2-SYNTROPHIN.	1.11e+00
8	56	52.8	635	14	O09496	K5 LYASE.	1.11e+00
9	56	52.8	1074	5	O94046	T13F2.3 PROTEIN.	1.11e+00
10	56	52.8	1533	4	O75560	MELASTATIN 1.	1.11e+00
11	56	52.8	3038	3	O9YGA5	LOVASTATIN NONAKETIDE	1.11e+00
12	56	52.8	3944	5	O18667	C47D12.1 PROTEIN.	1.11e+00
13	55	51.9	250	2	O9X117	METHIONINE AMINOPEPTID	1.85e+00
14	55	51.9	397	2	O33732	NITRATE REDUCTASE (EC	1.85e+00
15	55	51.9	475	5	O46038	EG-103B4.2 PROTEIN (1.85e+00
16	54	50.9	463	10	O39496	ALPHA 3 FRUSTULIN (FRA	3.08e+00
17	54	50.9	482	1	O27456	DNA POLYMERASE DELTA S	3.08e+00
18	54	50.9	682	11	O9WV16	573K1.1.3 (GAMMA-AMINO	3.08e+00
19	54	50.9	787	14	O56293	ORIGIN-BINDING PROTEIN	3.08e+00
20	54	50.9	812	11	O92308	GABAB RECEPTOR 1D.	3.08e+00

RA WOOD V., RAJANDREAM M.A., BARRELL B.G., TAYLOR K., HARRIS D.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031788; CAA21161.1;
 SQ SEQUENCE 157 AA; 17662 MW; FDD493DC CRC32;

Query Match 59.4%; Score 63; DB 3; Length 157;
 Best Local Similarity 61.5%; Pred. No. 2.53e-02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 90 QETKVKLVLCRG 102
 : | | | | | | | | | |
 QY 1 HEIKKVLVPGCHG 13

RESULT 3
 ID Q9ZU90 PRELIMINARY; PRT; 292 AA.
 AC Q9ZU90;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE T8011.21 PROTEIN.
 GN T8011.21

OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
 RA BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
 RA CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
 RA FRASER C.M., VENTER J.C.;
 RT "Arabidopsis thaliana chromosome II BAC T8011 genomic sequence."
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006069; AAD12709.1;
 SQ SEQUENCE 292 AA; 33005 MW; 9E18A63B CRC32;

Query Match 57.5%; Score 61; DB 10; Length 292;
 Best Local Similarity 63.6%; Pred. No. 7.66e-02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 109 ITRKIVPGCSG 119
 | | : | | | | | | |
 QY 3 IKKVLVPGCHG 13

RESULT 4
 ID Q31860 PRELIMINARY; PRT; 77 AA.
 AC Q31860;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE YQJC PROTEIN.
 GN YQJC.

OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RC MEDLINE; 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOVIN A., BORCHERT S.,
 RA BORRIS R., BOURSTIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLEICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPPI G., GUY B.J., HAGA K., HATCHE J., HARWOOD C.R., HENAUT A.,

RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGIMURA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHE E., ROCHE M., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCORFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis."
 RT Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z99114; CAB13841.1;
 SQ SEQUENCE 77 AA; 8518 MW; B25923A1 CRC32;

Query Match 53.8%; Score 57; DB 2; Length 77;
 Best Local Similarity 66.7%; Pred. No. 6.56e-01;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 14 RTVLVPACH 22
 : | | | | : | |
 QY 4 KKVLPVPGCH 12

RESULT 5
 ID Q45419 PRELIMINARY; PRT; 697 AA.
 AC Q45419;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.
 GN MTLR.

OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 7954;
 RC MEDLINE; 96421984.
 RA HENSTRA S.A., TOLNER B., HOEVE DUURKENS R.H., KONINGS W.N.,
 RA ROBILIARD G.T.;
 RT "Cloning, expression, and isolation of the mannitol transport protein
 RT from the thermophilic bacterium Bacillus stearothermophilus."
 RL J. Bacteriol. 178:5586-5591(1996).
 DR EMBL; U18943; AAC44464.1;
 DR PFAM; PF00874; BglG_antitermin; 2.
 SQ SEQUENCE 697 AA; 79242 MW; A78A4E30 CRC32;

Query Match 53.8%; Score 57; DB 2; Length 697;
 Best Local Similarity 58.3%; Pred. No. 6.56e-01;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 549 HSIKEVLAECACH 560
 | | | | | | | | | |
 QY 1 HEIKKVLVPGCH 12

RESULT 6 PRELIMINARY; PRT; 520 AA.
 ID Q61235

Q61235;
AC 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE SYNTROPHIN-2.
GN SNTB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE-DIAPHRAGM;
RX MEDLINE; 94000819.
RA ADAMS M.E., BUTLER M.H., DWYER T.M., PETERS M.F., MURNANE A.A.,
RA FROEHNER S.C.;
RT "Two forms of mouse syntrophin, a 58 kd dystrophin-associated protein,
RT differ in primary structure and tissue distribution.";
RL Neuron 11:531-540(1993).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE-DIAPHRAGM;
RX MEDLINE; 96029685.
RA ADAMS M.E., DWYER T.M., DOWLER L.L., WHITE R.A., FROEHNER S.C.;
RT "Mouse alpha 1- and beta 2-syntrophin gene structure, chromosome
RT localization, and homology with a discs large domain.";
RL J. Biol. Chem. 270:25859-25865(1995).
RL EMBL; U00678; AAC53060.1; -.
DR HSSP; Q12959; 1PDR.
DR MGD; MGI:101771; Sntb2.
DR PFAM; PF00595; PD2; 1.
SQ SEQUENCE 520 AA; 56381 MW; 6FC27580 CRC32;
Query Match 52.8%; Score 56; DB 11; Length 520;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 410 RILVQCHAA 419
QY :||| |||:
5 KVLVPGCHGS 14
RESULT 7
ID Q13425 PRELIMINARY; PRT; 540 AA.
AC Q13425;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BETA2-SYNTROPHIN.
GN SNT B2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RC MEDLINE; 96162017.
RA AHN A.H., FEENER C.A., GUSSONI E., YOSHIDA M., OZAWA E., KUNKEL L.M.;
RT "The three human syntrophin genes are expressed in diverse tissues,
RT have distinct chromosomal locations, and each bind to dystrophin and
RT its relatives.";
RL J. Biol. Chem. 271:2724-2730(1996).
RL EMBL; U40572; AAC50449.1; -.
DR HSSP; Q12959; 1PDR.
DR PFAM; PF00595; PD2; 1.
SQ SEQUENCE 540 AA; 57949 MW; 7016418A CRC32;
Query Match 52.8%; Score 56; DB 4; Length 540;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 430 RILVQCHAA 439
QY :||| |||:
5 KVLVPGCHGS 14
RESULT 8
ID Q09496 PRELIMINARY; PRT; 635 AA.
AC Q09496;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE K5 LYASE.
OS Coliphage K5.
OC Viruses.
RN [1]
RN SEQUENCE FROM N.A.
RA ESUMEH F., HANFLING P., JANN B., JANN K., ROBERTS I.S.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y10025; CAA71133.1; -.
KW Lyase.
SQ SEQUENCE 635 AA; 66944 MW; 58219280 CRC32;
Query Match 52.8%; Score 56; DB 14; Length 635;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 294 INNVLMSGCDGT 305
QY :||| |||:
3 IKVLVPGCHGS 14
RESULT 9
ID Q94046 PRELIMINARY; PRT; 1074 AA.
AC Q94046;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE T13F2.3 PROTEIN.
GN T13F2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RN SWINBURNE J.;
RN SEQUENCE FROM N.A.
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., HILLIER L., JIER M., JOHNSTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 281122; CA503354.1; -.
DR PFAM; PF00533; BRCT; 1.
SQ SEQUENCE 1074 AA; 120339 MW; DD8BB51C CRC32;
Query Match 52.8%; Score 56; DB 5; Length 1074;
Best Local Similarity 42.9%; Pred. No. 1.11e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 565 HQITHVLVDSRCRT 578
QY :||| |||:
1 HEIKKVLVPGCHGS 14
RESULT 10
ID Q75560 PRELIMINARY; PRT; 1533 AA.
AC Q75560;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BETA2-SYNTROPHIN.
GN SNT B2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RC MEDLINE; 96162017.
RA AHN A.H., FEENER C.A., GUSSONI E., YOSHIDA M., OZAWA E., KUNKEL L.M.;
RT "The three human syntrophin genes are expressed in diverse tissues,
RT have distinct chromosomal locations, and each bind to dystrophin and
RT its relatives.";
RL J. Biol. Chem. 271:2724-2730(1996).
RL EMBL; U40572; AAC50449.1; -.
DR HSSP; Q12959; 1PDR.
DR PFAM; PF00595; PD2; 1.
SQ SEQUENCE 540 AA; 57949 MW; 7016418A CRC32;
Query Match 52.8%; Score 56; DB 4; Length 540;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 430 RILVQCHAA 439
QY :||| |||:
5 KVLVPGCHGS 14

AC 075560; (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE MELASTATIN 1.
 GN MSLN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-RETINA;
 RA HUNTER J.J., SHAO J., SMUTKO J.S., DUSSAULT B.J., NAGLE D.L.,
 RA WOOLF E.A., HOLMGREN L.M., MOORE K.J., SHYJAN A.W.;
 RT "Chromosomal localization and genomic characterization of the mouse
 RT melastatin gene (Msln1).";
 RL melastatin (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF071787; AAC8000.1; -
 SQ SEQUENCE 1533 AA; 174412 MW; E7693DD2 CRC32;
 Query Match 52.8%; Score 56; DB 4; Length 1533;
 Best Local Similarity 46.2%; Pred. No. 1.11e+00;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Db 1241 DVKTHLVPCEONS 1253
 QY 2 EIRKVLPGCHGS 14
 RESULT 11
 ID Q9Y8A5 PRELIMINARY; PRT; 3038 AA.
 AC Q9Y8A5;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE LOVASTATIN NONKETIDE SYNTHASE.
 GN LOVB.
 OS Aspergillus terreus.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 20542;
 RX MEDLINE; 99310964.
 RA HENDRICKSON L., DAVIS C.R., ROACH C., NGUYEN D.K., ALDRICH T.,
 RA MCADA P.C., REEVES C.D.;
 RT "lovastatin biosynthesis in Aspergillus terreus: characterization of
 RT blocked mutants, enzyme activities and a multifunctional polyketide
 RT synthase gene.";
 RL Chem. Biol. 6:429-439(1999).
 DR EMBL: AF151722; AAD39830.1; -
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 KW Transferase.
 SQ SEQUENCE 3038 AA; 335001 MW; 90A5331A CRC32;
 Query Match 52.8%; Score 56; DB 3; Length 3038;
 Best Local Similarity 54.5%; Pred. No. 1.11e+00;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 2046 QRMKVLMPGC 2056
 QY 1 HEIRKVLVPGC 11
 RESULT 12
 ID Q18667 PRELIMINARY; PRT; 3944 AA.
 AC Q18667;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)
 DE C47D12.1 PROTEIN.
 GN C47D12.1.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GAJADSTY S.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z69902; CAA93765.1; -
 SQ SEQUENCE 3944 AA; 452752 MW; C44B1EB5 CRC32;
 Query Match 52.8%; Score 56; DB 5; Length 3944;
 Best Local Similarity 50.0%; Pred. No. 1.11e+00;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 1991 VKKILIECH 2000
 QY 3 IKKVLVPGCH 12
 RESULT 13
 ID Q9X1I7 PRELIMINARY; PRT; 250 AA.
 AC Q9X1I7;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP).
 GN TM1478.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99287316.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
 RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RT "Evidence for lateral gene transfer between Archaea and bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 395:323-329(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
 RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 CC PROTEINS...
 CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +
 CC PEPTIDE.
 CC -1- COFACTOR: COBALT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24A; ALSO KNOWN AS THE
 CC MAP FAMILY.
 DR EMBL; AE001798; AAD36544.1; -

DR PROSITE; PS00680; MAP_1; 1.
 KW Aminopeptidase; Hydrolase; Cobalt.
 SQ SEQUENCE 250 AA; 27489 MW; BA9A64CF CRC32;

Query Match 51.9%; Score 55; DB 2; Length 250;
 Best Local Similarity 60.0%; Pred. No. 1.85e+00;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 25 REVRKVIIVPG 34
 QY 1 HEIKKVLVPG 10

RESULT 14
 ID O33732 PRELIMINARY; PRT; 397 AA.

AC O33732;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-JAN-1998 (TEMBLrel. 05, Last annotation update)
 DE NITRATE REDUCTASE (EC 1.7.99.4) (FRAGMENT).
 OS Shewanella putrefaciens.
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 OC Shewanella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCIMB400;
 RA GORDON E.H.J., PIKE A.D., FISCHER H., CHAPMAN S.K., REID G.A.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ000006; CAA03851.1;
 KW Oxidoreductase.
 FT NON_TER 397
 SQ SEQUENCE 397 AA; 42187 MW; 5F0AF4AC CRC32;

Query Match 51.9%; Score 55; DB 2; Length 397;
 Best Local Similarity 60.0%; Pred. No. 1.85e+00;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 290 KVLVATSCQG 299
 QY 4 KVLVPGCHG 13

RESULT 15
 ID O46038 PRELIMINARY; PRT; 475 AA.

AC O46038;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE EG:103B4.2 PROTEIN.
 GN EG:103B4.2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MURPHY L., HARRIS D., BARRELL B.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BENOS P.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL009193; CAA15692.1;
 DR FLYBASE; FBgn0023550; EG:103B4.2.
 SQ SEQUENCE 475 AA; 53756 MW; FBFF0FEE CRC32;

Query Match 51.9%; Score 55; DB 5; Length 475;
 Best Local Similarity 55.6%; Pred. No. 1.85e+00;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 343 KNLMPGCH 351
 QY 4 KVLVPGCH 12

Search completed: Sun Apr 2 17:54:11 2000
 Job time : 70 secs.

MPERCH_PP

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:58:10 2000; MasPar time 75.60 Seconds
125.641 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-3
Description: (1-137) from US09362731.pep
Perfect score: 1057
Sequence: 1 DQYKANSKFTIGTELGGQY.....FGCHGSEPCIIHRGKPF SR 137

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb112

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp_mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp_unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 34.307; Variance 56.239; scale 0.610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	10.0	145	5	GROUP 2 ALLERGEN EUR M	1.65e-05
2	94	8.9	1792	13	LAMININ BETA 2-LIKE CH	2.41e-03
3	90	8.5	1000	2	CATION EFFLUX (ACRB/AC	1.19e-02
4	89	8.4	3871	5	ZC116.3 PROTEIN.	1.76e-02
5	88	8.3	313	5	EXTRACELLULAR MATRIX P	2.60e-02
6	88	8.3	379	5	GLUE PROTEIN.	2.60e-02
7	87	8.2	988	6	GASTRIC MUCIN (FRAGMEN	3.83e-02
8	83	7.9	373	5	FMRFAMIDE (FRAGMENT).	1.77e-01
9	83	7.9	595	10	NEOXANTHIN CLEAVAGE EN	1.21e-01
10	84	7.9	3623	11	INTRINSIC FACTOR-B12 R	2.58e-01
11	82	7.8	127	5	FMRFAMIDE (FRAGMENT).	2.58e-01
12	82	7.8	925	14	POSS. PRECURSOR POLYPE	2.58e-01
13	82	7.8	929	14	POSS. PRECURSOR POLYPE	2.58e-01
14	82	7.8	1193	13	C-SERATE-1 PROTEIN (FR	2.58e-01
15	81	7.7	177	5	ORF-REF1.	3.74e-01
16	81	7.7	3623	4	INTRINSIC FACTOR-B12 R	3.74e-01
17	80	7.6	356	6	UNKNOWN MRNA, PARTIAL	5.41e-01
18	79	7.5	1081	4	MUC5AC PROTEIN (FRAGME	7.82e-01
19	78	7.4	363	14	THYMIDINE KINASE.	1.13e+00
20	78	7.4	752	13	NOTCH RECEPTOR PROTEIN	1.13e+00

21	78	7.4	1042	4	Q13792	1.13e+00
22	78	7.4	1219	11	Q63722	1.13e+00
23	78	7.4	2704	5	Q97458	1.13e+00
24	77	7.3	199	2	Q92E53	1.62e+00
25	77	7.3	398	2	Q9X7N6	1.62e+00
26	77	7.3	750	4	Q08424	1.62e+00
27	77	7.3	1247	2	Q07910	1.62e+00
28	77	7.3	1764	5	Q02650	1.62e+00
29	77	7.3	4289	4	Q78530	1.62e+00
30	76	7.2	107	10	Q24233	1.62e+00
31	76	7.2	107	10	Q24232	2.31e+00
32	76	7.2	154	10	Q39814	2.31e+00
33	76	7.2	187	2	Q9X5H4	2.31e+00
34	76	7.2	421	3	Q13960	2.31e+00
35	76	7.2	443	2	Q92422	2.31e+00
36	76	7.2	599	4	Q9Y2N9	2.31e+00
37	76	7.2	894	11	Q88715	2.31e+00
38	75	7.1	128	5	Q61610	3.30e+00
39	75	7.1	263	4	Q99740	3.30e+00
40	75	7.1	280	14	Q89642	3.30e+00
41	75	7.1	1218	4	Q15816	3.30e+00
42	75	7.1	1218	4	Q14902	3.30e+00
43	75	7.1	1218	4	Q15122	3.30e+00
44	75	7.1	1227	4	Q78504	3.30e+00
45	75	7.1	2026	4	Q00468	3.30e+00

ALIGNMENTS

RESULT 1
ID O96430 PRELIMINARY; PRT; 145 AA.
AC O96430;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Pyroglyphidae; Pyroglyphidae;
OC Eukaryophus.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH W., HART B.J., THOMAS W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047613; AAC82349.1;
DR HSP; P49278; IAGV.
SQ SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;

Query Match 10.0%; Score 106; DB 5; Length 145;
Best Local Similarity 75.0%; Pred. No. 1.65e-05;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 36 GCGSEPCVHRGTA 51
QY 52 GCHGSEPCIIHRGKPF 67
|| |||||:||||:|

RESULT 2
ID O57484 PRELIMINARY; PRT; 1792 AA.
AC O57484;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE LAMININ BETA 2-LIKE CHAIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93015947.
RA O'REAR J.J.;
RT "A novel laminin B1 chain variant in avian eye."

RL J. Biol. Chem. 267:20555-20557(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA LIU J., SWADISON S., XIE W., BREWTON R.G., MAYNE R.;
 RL Matrix Biol. 16:0-0(1998).
 DR EMBL; AF038555; AAB92586.1; -.
 DR HSSP; P02468; IKLO.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
 DR PFAM; PF00053; laminin_EGF; 13.
 DR PFAM; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 KW Glycoprotein.
 SQ SEQUENCE 1792 AA; 195722 MW; AFPC9020 CRC32;
 Query Match 8.9%; Score 94; DB 13; Length 1792;
 Best Local Similarity 31.7%; Pred. No. 2.41e-03;
 Matches 26; Conservative 15; Mismatches 32; Indels 9; Gaps 8;
 Db 1080 LGSQGCPCACHPQHSLSAPACNFTQCSCRPQ--FGG-RTCANCOEQHWGDPRLQCR A 1136
 QY 50 FGCGHSGEPCIIHRGKPFSS-SCHG-SEPCIIHRGKPFSGCHGSEPCII-HRGKPFSSCHG 106
 Db 1137 CD-CD-PRGIASCTCHRGSGHC 1156
 QY 107 SEPCIIHRGKPFSGCH-GSEPC 127
 RESULT 3
 ID 066977 PRELIMINARY; PRT; 1000 AA.
 AC 066977;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE CATION EFFLUX (ACRB/ACRD/ACRF FAMILY).
 GN ACRD4.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE; 98196666.
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus".
 RT Nature 392:353-358(1998).
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE000707; AAC06934.1; -.
 DR PFAM; PF00873; ACR_tran; 1.
 DR PRINTS; PR00702; ACRIFLAVINRP.
 SQ SEQUENCE 1000 AA; 112872 MW; AA50FB98 CRC32;
 Query Match 8.5%; Score 90; DB 2; Length 1000;
 Best Local Similarity 48.5%; Pred. No. 1.19e-02;
 Matches 16; Conservative 6; Mismatches 9; Indels 2; Gaps 2;
 Db 739 YKAEEDFVKIHNLRVYIPARGELIPITEL 771
 QY 3 YIKANSKFI-GITELGGQYIKA-NSKFIGITEL 33
 RESULT 4
 ID Q20911 PRELIMINARY; PRT; 3871 AA.
 AC Q20911; Q23242;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE ZC116.3 PROTEIN.
 GN ZC116.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BURTON J.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans".
 RT Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA SMYE R.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z74473; CAA98952.1; -.
 DR EMBL; Z74046; CAA98952.1; JOINED.
 DR EMBL; Z74046; CAA98957.1; -.
 DR EMBL; Z74473; CAA98957.1; JOINED.
 DR HSSP; P00740; IIXA.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR PFAM; PF00431; CUB; 13.
 DR PFAM; PF00008; EGF; 7.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 3871 AA; 433659 MW; OB3CAF62 CRC32;
 Query Match 8.4%; Score 89; DB 5; Length 3871;
 Best Local Similarity 22.6%; Pred. No. 1.76e-02;
 Matches 21; Conservative 20; Mismatches 47; Indels 5; Gaps 5;
 Db 56 GLLVKMFRQIIIGINKLIAINACDPNKCNSGGTCIPFCAKFTCLCPHPHTGTTCADIDE 115
 QY 18 GOYIKANSKFIGITELSSCHGSEPCIIHRGKPFSGCHGSE-PCIIHRGKPFSSCHGS-EP 75
 Db 116 CSVINGTT-AGCQNGTICINNRG-GFECQCSG 146
 QY 76 CIIHRGKPFSGCHGSEPCIIHRGKPFSS-CHGS 107
 RESULT 5
 ID Q24330 PRELIMINARY; PRT; 313 AA.
 AC Q24330;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE EXTRACELLULAR MATRIX PROTEIN B (FRAGMENT).
 GN ECMB.
 OS Dictyostelium minutum.
 OC Eukaryota; Dictyosteliida; Dictyostelium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-71-2;
 RX MEDLINE; 94336717.
 RA VAN ES S., NIEUWENHUIJSEN B.W., LENOUEVEL F., VAN DEURSEN E.M.,
 RA SCHAAP P.;
 RT "Universal signals control slime mold stalk formation".
 RL proc. Natl. Acad. Sci. U.S.A. 91:8219-8223(1994).
 DR EMBL; X78948; CAA55545.1; -.

```

DR EMBL; Z29565; CAA82671.1; -.
DR EMBL; X76203; CAA53797.1; -.
DR FLYBASE; FBgn0C11269; Dvir\Lgp3
FT DOMAIN 36 39 POLY-THR.
SQ SEQUENCE 379 AA; 41083 MW; D832E6B3 CRC32;

Query Match      8.38; Score 88; DB 5; Length 379;
Best Local Similarity 24.08; Pred. No. 2.60e-02;
Matches 23; Conservative 22; Mismatches 46; Indels 5; Gaps 5;

Db 80 TKPTTTRTKP-TTTRTKTPTTTRTKP-TTTRTKPTTTRTKP-TTTRTKPTTTRRT 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 39 SEPCIIHRGPFGGCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFGGCHGSEPCIHRG 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 KP-TTTRTKTPTTTRTKP-TTTRTKTPTTTRTKP 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 KPFSSCHGSEPCIIHRGKPFGGCHGSEPCIIHRGK 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID O97867 PRELIMINARY; PRT; 988 AA.
AC O97867;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE DE GASTRIC MUCIN (FRAGMENT).
GN MUC5AC.
OS Sus scrofa domestica (domestic pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[]]
RP SEQUENCE OF 35-109 FROM N.A.
RC TISSUE-GASTRIC EPITHELIUM;
RX MEDLINE: 95275264.
RA TURNER B.S., BHASKAR K.R., HADZOPOULOU-CLADARAS M., SPECIAN R.D.,
RA LAMONT J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-GASTRIC EPITHELIUM;
RA TURNER B.S., BHASKAR K.R., HADZOPOULOU-CLADARAS M., LAMONT J.T.;
RT "Cysteine-rich regions of pig gastric mucin contain cystine knot and
RT von Willebrand factor domains at the 3'-carboxyl terminal: similarity
RT to human and rat MUC5AC."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF054584; AAD19833.1; -.
FT NON_TER 1.
SQ SEQUENCE 988 AA; 105264 MW; 2D3E3F11 CRC32;

Query Match      8.28; Score 87; DB 6; Length 988;
Best Local Similarity 22.28; Pred. No. 3.83e-02;
Matches 22; Conservative 25; Mismatches 47; Indels 5; Gaps 4;

Db 452 VTLPGOPPCHAPPRLTWTVEPTPTTSCPPSPICQLILSEVPACHAETPPWPFQGVCF 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 28 IGTEYSSCHGSEP-CIHRGKPFGGCHGSEPCIIHRGKPFSSCHGSEPC-IIHRGKPF 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 DHCHMPDVLCSGLELYAALCASLGVCIIDWRGRTNHTC 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 85 GGCHGSEPCIIHRG-KPFSS-CHGSEPCIIHRGKPFGGC 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID Q16964 PRELIMINARY; PRT; 373 AA.
AC Q16964;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE FMRFAMIDE (FRAGMENT).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anasoidae;

```

[illegible]

```
RC TISSUE-SPINAL CORD;
RX MEDLINE; 96175595.
RA MYAT A.; HENRIQUE D., ISH-HOROWICZ D., LEWIS J.;
RT "A chick homologue of Serrate and its relationship with Notch and
RL Delta homologues during central neurogenesis.";
RD Dev. Biol. 174:233-247(1996).
DR EMBL; X95283; CAA64604.1; -.
DS HSSP; P00740; IIXA.
DR PROSITE; PS00010; ASX_HYDROXYL; 10.
DR PROSITE; PS01167; BGF_CA; 8.
DR PFAM; PF01414; DSL; 1.
DR PFAM; PF00008; EGF; 14.
DR Glycoprotein; EGF-like domain.
FT NON_TER      1
SQ SEQUENCE     1193 AA; 131039 MW; 55E5FCD1 CRC32;

Query Match       7.8%; Score 82; DB 13; Length 1193;
Best Local Similarity 30.2%; Pred.No. 2.58e-01;
Matches    13; Conservative   10; Mismatches    16; Indels    4; Gaps    4;

Db 871 C-GPRPCIIH-AKGNCCPAGHACVPVKEDHCFTHPCAAAGEC 911
Qy 36 CHGSEPCIHRGPFGCGHGSEPCI-IHRGKPFS-SCHGSEPC 76

RESULT 15
ID QJ6577 PRELIMINARY; PRT; 177 AA.
AC QJ6577;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DE DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE ORF-RFL.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae;
OC Schistosomatidae; Schistosomatiidae; Schistosome.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86287295.
RA BOBEK L.; REKOSH D.M.; VAN KEULEN H.; LOVERDE P.T.;
RT "Characterization of a female-specific cDNA derived from a
developmentally regulated mRNA in the human blood fluke Schistosoma
mansoni.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5544-5548(1986).
DR EMBL; M14309; AAA74695.1; -.
DS HSSP; P30129; 4DPV.
SQ SEQUENCE 177 AA; 16332 MW; F17CF2EF CRC32;

Query Match       7.7%; Score 81; DB 5; Length 177;
Best Local Similarity 32.1%; Pred.No. 3.74e-01;
Matches    26; Conservative   18; Mismatches    26; Indels    11; Gaps    11;

Db 31 GGCVGS-DMSGVDSDSGGGCTGSD-CGGGYGGYGCGCSGD-CGNY-GGGYGDCNG 86
Qy 51 GGCHGESEPCIHRGKP-FSS-CHGSEPCIHRGKPFEG-CHGSEPCIHRGKPFSS-CHG 106
Db 87 GD-CGNY-GGGYGNGGG--C 104
Qy 107 SEPCIHRGKPFEGCHGSEPC 127

Search completed: Sun Apr 2 17:59:30 2000
Job time : 80 secs.
```

Search completed: Sun Apr 2 17:59:30 2000
Job time : 80 secs.

MPSEARCH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:57:16 2000; MasPar time 30.92 Seconds
Tabular output not generated. 132.305 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.ppe
Perfect Score: 1057
Sequence: 1 DOYKANSKFIGITELGGOY.....FGGCHGSEPCIIHRGKPFPSR 137

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 35.712; Variance 57.531; scale 0.621

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	133	12.6	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	6.73e-11
2	132	12.5	146	1	DER2_DERPT MITE ALLERGEN DER P 2	1.07e-10
3	104	9.8	1314	1	TETX_CLOTE TETANUS TOXIN PRECURSO	2.67e-05
4	91	8.6	647	1	ZG48_XENLA GASTRULA ZINC FINGER P	5.41e-03
5	89	8.4	1798	1	LMB2_HUMAN LAMININ BETA-2 CHAIN P	1.18e-02
6	88	8.3	713	1	DDX4_RAT DEAD BOX PROTEIN 4 (VA	1.74e-02
7	86	8.1	552	1	FMR2_APLCA FMRFAMIDE NEUROPEPTIDE	3.76e-02
8	85	8.0	435	1	2022_XENLA OOCYTE ZINC FINGER PRO	5.50e-02
9	85	8.0	1046	1	PSTA_DICDI PRESTALK PROTEIN PRECU	5.50e-02
10	85	8.0	1801	1	LMB2_RAT LAMININ BETA-2 CHAIN P	5.50e-02
11	84	7.9	309	1	FMR1_APLCA FMRFAMIDE NEUROPEPTIDE	8.02e-02
12	82	7.8	1046	1	VGNM_CPMV GENOME POLYPROTEIN M	1.69e-01
13	79	7.5	2703	1	NOTC_DROME NEUROGENIC LOCUS NOTCH	5.08e-01
14	78	7.4	94	1	SAPE_SARPE SAPECN PRECURSOR	7.27e-01
15	78	7.4	94	1	DEFI_PROTE PHORMICIN PRECURSOR	7.27e-01
16	78	7.4	177	1	EGG3_SCHWA EGGSHELL PROTEIN PRECU	7.27e-01
17	78	7.4	177	1	EGG2_SCHWA EGGSHELL PROTEIN PRECU	7.27e-01
18	78	7.4	364	1	KITH_ILTVT THYMIDINE KINASE (EC 2	7.27e-01
19	78	7.4	637	1	DDX4_MOUSE DEAD BOX PROTEIN 4 (VA	7.27e-01
20	78	7.4	732	1	ACPH_HUMAN ACYLAMINO-ACID-RELEASE	1.04e-00
21	77	7.3	261	1	YH86_SYN3 HYPOTHETICAL 29.3 KD P	1.04e-00
22	77	7.3	757	1	COMP_HUMAN CARTILAGE OLIGOMERIC M	1.04e-00
23	76	7.2	755	1	COMP_RAT CARTILAGE OLIGOMERIC M	1.48e-00

24	76	7.2	1799	1	LMB2_MOUSE LAMININ BETA-2 CHAIN P	1.48e+00
25	75	7.1	141	1	ALG2_TYRPU GROUP 2 ALLERGEN PRECU	2.10e+00
26	74	7.0	1147	1	CGA1_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
27	74	7.0	1182	1	CGA2_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
28	74	7.0	1186	1	CGA0_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
29	73	6.9	362	1	PHOSPHO-2-DEHYDRO-3-DE	4.19e+00
30	73	6.9	1056	1	MUC5_HUMAN TRACHEOBRONCHIAL MUCIN	4.19e+00
31	73	6.9	1064	1	SYIC_SCHPO ISOLEUCYL-TRNA SYNTHET	4.19e+00
32	73	6.9	1713	1	LMA3_HUMAN LAMININ ALPHA-3 CHAIN	4.19e+00
33	73	6.9	4660	1	LRP2_RAT LOW-DENSITY LIPOPROTEI	4.19e+00
34	72	6.8	141	1	LEP1_LEPDS MITE ALLERGEN LEP D 1	5.88e+00
35	72	6.8	216	1	RL3_RICPR 50S RIBOSOMAL PROTEIN	5.88e+00
36	72	6.8	245	1	MTA_METJA TETRAHYDROMETHANOPTERI	5.88e+00
37	72	6.8	305	1	ACOC_SOLTI ACYCLAMINO-ACID-RELEASE	5.88e+00
38	72	6.8	616	1	ACPH_RAT ACYLAMINO-ACID-RELEASE	5.88e+00
39	72	6.8	732	1	MSPI_PLAFK MEROZOITE SURFACE PROT	5.88e+00
40	72	6.8	1630	1	MSPI_PLAFM MEROZOITE SURFACE PROT	5.88e+00
41	72	6.8	1639	1	NTC4_MOUSE NEUROGENIC LOCUS NOTCH	5.88e+00
42	72	6.8	1964	1	LMA1_HUMAN LAMININ ALPHA-1 CHAIN	5.88e+00
43	72	6.8	3075	1	FMR4_APLCA FMRFAMIDE NEUROPEPTIDE	8.22e+00
44	71	6.7	196	1	FMR4_APLCA FMRFAMIDE NEUROPEPTIDE	8.22e+00
45	71	6.7	488	1	ZF92_MOUSE ZINC FINGER PROTEIN 92	8.22e+00

ALIGNMENTS

RESULT 1
ID DEF2_DERFA STANDARD; PRT; 146 AA.
AC Q00855; P39672; Q26359;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
GN DERF2.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 91291341.
RA YUUKI T., OKUMURA Y., ANDO T., YAMAKAWA H., SUKO M., HAYDA M.,
OKUDAIRA H.;
RT "Cloning and expression of cDNA coding for the major house dust mite
allergen Der f II in Escherichia coli.";
RL Agric. Biol. Chem. 55:1233-1238(1991).
RN [2]
RP SEQUENCE OF 4-146 FROM N.A.
RX MEDLINE; 94256850.
RA OKUHIRA H.;
RT "Molecular biology of mite antigens.";
RL Aterugi 43:435-440(1994).
RN [3]
RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
RX MEDLINE; 93283958.
RA NISHIYAMA C., YUUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
RT "Determination of three disulfide bonds in a major house dust mite
allergen, Der f II.";
RL Int. Arch. Allergy Immunol. 101:159-166(1993).
RN [4]
RP PARTIAL SEQUENCE OF 18-52.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
PLATTS-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE; 98079068.
RA ICHIKAWA S., HATANAKA H., YUUKI T., IWAMOTO N., KOJIMA S.,
NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
RT "Solution structure of Der f 2, the major mite allergen for atopic

diseases.";
 J. Biol. Chem. 273:356-360(1998).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10447; BAA01239.1; -;
 DR EMBL; D10448; BAA01240.1; -;
 DR EMBL; D10449; BAA01241.1; -;
 DR EMBL; S70378; AAB30829.1; -;
 DR PIR; PS0417; PS0417.
 DR PDB; 1AHK; 08-APR-98.
 DR PDB; 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT VARIANT 90 95
 FT VARIANT 105 105 M -> V (IN CLONE 1).
 FT VARIANT 105 105 I -> A (IN CLONE 11).
 FT VARIANT 128 128 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
 SQ SEQUENCE 146 AA; 15802 MW; 72623F23 CRC32;
 Query Match 12.6%; Score 133; DB 1; Length 146;
 Best Local Similarity 88.2%; Pred. No. 6.73e-11;
 Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 37 GCHGSEPCIIHRGKPF 53
 QY 52 GCHGSEPCIIHRGKPF 68
 RESULT 2
 ID DER2_DERPT STANDARD; PRT; 146 AA.
 AC P49278;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
 GN DERP2.
 OS Dermatophagoides pteronyssinus (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
 OC Dermatophagoides.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
 RA THOMAS W.R.;
 RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
 RT plaque immunoassay";
 RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
 RN [2]
 RP PARTIAL SEQUENCE OF 18-57.
 RX MEDLINE; 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATTS-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [3]

RP STRUCTURE BY NMR.
 RX MEDLINE; 98409423.
 RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
 RT "Tertiary structure of the major house dust mite allergen Der p 2:
 RT sequential and structural homologies.";
 RL Biochemistry 37:12707-12714(1998).
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC PDB; 1A9V; 14-OCT-98.
 DR Allergen; Signal; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER P 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;
 Query Match 12.5%; Score 132; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 1.07e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 37 GCHGSEPCIIHRGKPF 52
 QY 52 GCHGSEPCIIHRGKPF 67
 RESULT 3
 ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: Primary structure, expression in E. coli, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911;
 RX MEDLINE; 87040747.
 RA FAIRWEATHER N.F., LYNESS V.A.;
 RT "The complete nucleotide sequence of tetanus toxin.";
 RL Nucleic Acids Res. 14:7809-7812(1986).
 RN [3]
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE; 86085672.
 RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in Escherichia coli.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE; 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE; 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]

RP IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE: 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE: 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL Nature 359:832-835(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE: 97475217.
 RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL Nat. Struct. Biol. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMALIN-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMALIN.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
 CC GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04436; CAA28033.1; -.
 DR EMBL: M12739; AAA23282.1; -.
 DR EMBL: X06214; CAA29564.1; -.
 DR PIR: A25689; BTCLTN.
 DR PDB: 1AF9; 29-APR-98.
 DR PDB: 1A8D; 14-OCT-98.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
 FT 3D-structure. 0 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 233 233 BY SIMILARITY.
 FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT DISULFID 438 466 INTERCHAIN.
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
 Query Match 9.88; Score 104; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 2.67e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYIKANSKFIGITEL 843
 QY 2 QYIKANSKFIGITEL 16
 RESULT 4
 ID ZG48_XENLA STANDARD; PRT; 647 AA.
 AC P18723;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE GASTRULA ZINC FINGER PROTEIN XLCGF48.2 (FRAGMENT).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE OF 1-339 FROM N.A.
 RX MEDLINE: 89345612.
 RA KNOCHHEL W., POTING A., KOSTER M., EL BARADI T., NIETZELD W.,
 RA BOUMWELSTER T., PIELER T.;
 RT "Evolutionary conserved modules associated with zinc fingers in
 RT Xenopus laevis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
 RN [2]
 RP SEQUENCE OF 312-647 FROM N.A.
 RX MEDLINE: 90040698.
 RA NIETZELD W., EL-BARADI T., MENTZEL H., PIELER T., KOSTER M.,
 RA POTING A., KNOCHHEL W.;
 RT "Second-order repeats in Xenopus laevis finger proteins.";
 RL J. Mol. Biol. 208:639-659(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M25871; AAA50018.1; -.
 DR PIR: F33282; F33282.
 DR PIR: S06573; S06573.
 DR HSP: P08047; LSP2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2; 12.
 DR PFAM: PF00096; zf-C2H2; 12.
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat.
 FT ZN_FING 317 339 C2H2-TYPE.
 FT ZN_FING 345 367 C2H2-TYPE.
 FT ZN_FING 373 395 C2H2-TYPE.
 FT ZN_FING 401 423 C2H2-TYPE.
 FT ZN_FING 429 451 C2H2-TYPE.
 FT ZN_FING 457 479 C2H2-TYPE.
 FT ZN_FING 485 507 C2H2-TYPE.
 FT ZN_FING 513 535 C2H2-TYPE.
 FT ZN_FING 540 563 C2H2-TYPE.
 FT ZN_FING 541 563 C2H2-TYPE.
 FT ZN_FING 569 591 C2H2-TYPE.
 FT ZN_FING 597 619 C2H2-TYPE.
 FT ZN_FING 625 647 C2H2-TYPE.
 FT NON_TER 647 647
 SQ SEQUENCE 647 AA; 73054 MW; C969B307 CRC32;
 Query Match 8.6%; Score 91; DB 1; Length 647;
 Best Local Similarity 36.7%; Pred. No. 5.41e-03;
 Matches 22; Conservative 9; Mismatches 24; Indels 5; Gaps 5;
 Db 412 KFSRLTEHRCHTGEFFVCSECKSPAQSPQLDHLRIHTGEPFCS-CSDGKCFORA 470
 QY 26 KFTGITELSSCH-GSEPCI-IHRGKPGGCHGSEPCI-IHRG-KPFSSCHGSEPCIHRG 81

FT	DOMAIN	283	346	V).
FT	DOMAIN	347	409	LAMININ EGF-LIKE 1.
FT	DOMAIN	347	409	LAMININ EGF-LIKE 2.
FT	DOMAIN	410	469	LAMININ EGF-LIKE 3.
FT	DOMAIN	470	521	LAMININ EGF-LIKE 4.
FT	DOMAIN	522	552	LAMININ EGF-LIKE 5 (INCOMPLETE).
FT	DOMAIN	553	781	LAMININ DOMAIN IV.
FT	DOMAIN	782	1190	8 X LAMININ EGF-LIKE REPEATS (DOMAIN I II).
FT	DOMAIN	783	830	LAMININ EGF-LIKE 6.
FT	DOMAIN	831	876	LAMININ EGF-LIKE 7.
FT	DOMAIN	877	926	LAMININ EGF-LIKE 8.
FT	DOMAIN	927	985	LAMININ EGF-LIKE 9.
FT	DOMAIN	986	1037	LAMININ EGF-LIKE 10.
FT	DOMAIN	1038	1094	LAMININ EGF-LIKE 11.
FT	DOMAIN	1095	1142	LAMININ EGF-LIKE 12.
FT	DOMAIN	1143	1189	LAMININ EGF-LIKE 13.
FT	DOMAIN	1190	1409	DOMAIN II.
FT	DOMAIN	1410	1442	DOMAIN ALPHA.
FT	DOMAIN	1443	1798	DOMAIN I.
FT	DOMAIN	1453	1319	COILED COIL (POTENTIAL).
FT	DOMAIN	1472	1526	COILED COIL (POTENTIAL).
FT	DOMAIN	1577	1790	BY SIMILARITY.
FT	DISULFID	283	292	BY SIMILARITY.
FT	DISULFID	285	310	BY SIMILARITY.
FT	DISULFID	312	321	BY SIMILARITY.
FT	DISULFID	324	344	BY SIMILARITY.
FT	DISULFID	347	356	BY SIMILARITY.
FT	DISULFID	349	374	BY SIMILARITY.
FT	DISULFID	377	386	BY SIMILARITY.
FT	DISULFID	389	407	BY SIMILARITY.
FT	DISULFID	410	423	BY SIMILARITY.
FT	DISULFID	412	438	BY SIMILARITY.
FT	DISULFID	440	449	BY SIMILARITY.
FT	DISULFID	452	467	BY SIMILARITY.
FT	DISULFID	470	484	BY SIMILARITY.
FT	DISULFID	472	491	BY SIMILARITY.
FT	DISULFID	493	502	BY SIMILARITY.
FT	DISULFID	505	519	BY SIMILARITY.
FT	DISULFID	783	795	BY SIMILARITY.
FT	DISULFID	785	802	BY SIMILARITY.
FT	DISULFID	804	813	BY SIMILARITY.
FT	DISULFID	816	828	BY SIMILARITY.
FT	DISULFID	831	843	BY SIMILARITY.
FT	DISULFID	833	850	BY SIMILARITY.
FT	DISULFID	852	861	BY SIMILARITY.
FT	DISULFID	864	874	BY SIMILARITY.
FT	DISULFID	877	886	BY SIMILARITY.
FT	DISULFID	879	893	BY SIMILARITY.
FT	DISULFID	896	905	BY SIMILARITY.
FT	DISULFID	908	924	BY SIMILARITY.
FT	DISULFID	927	943	BY SIMILARITY.
FT	DISULFID	929	954	BY SIMILARITY.
FT	DISULFID	956	965	BY SIMILARITY.
FT	DISULFID	968	983	BY SIMILARITY.
FT	DISULFID	986	1000	BY SIMILARITY.
FT	DISULFID	988	1007	BY SIMILARITY.
FT	DISULFID	1010	1019	BY SIMILARITY.
FT	DISULFID	1022	1035	BY SIMILARITY.
FT	DISULFID	1095	1107	BY SIMILARITY.
FT	DISULFID	1097	1114	BY SIMILARITY.
FT	DISULFID	1116	1125	BY SIMILARITY.
FT	DISULFID	1128	1140	BY SIMILARITY.
FT	DISULFID	1143	1155	BY SIMILARITY.
FT	DISULFID	1145	1162	BY SIMILARITY.
FT	DISULFID	1164	1173	BY SIMILARITY.
FT	DISULFID	1176	1187	BY SIMILARITY.
FT	DISULFID	1190	1190	INTERCHAIN (PROBABLE).
FT	DISULFID	1193	1193	INTERCHAIN (PROBABLE).
FT	DISULFID	1797	1797	INTERCHAIN (PROBABLE).
FT	CARBOHYD	248	248	POTENTIAL.
FT	CARBOHYD	368	368	POTENTIAL.
FT	CARBOHYD	1085	1085	POTENTIAL.

FT MOD_RES 118 AMIDATION (G-119 PROVIDE AMIDE GROUP).
FT MOD_RES 156 AMIDATION (G-157 PROVIDE AMIDE GROUP).
FT MOD_RES 163 AMIDATION (G-164 PROVIDE AMIDE GROUP).
FT MOD_RES 178 AMIDATION (G-179 PROVIDE AMIDE GROUP).
FT MOD_RES 194 AMIDATION (G-195 PROVIDE AMIDE GROUP).
FT MOD_RES 201 AMIDATION (G-202 PROVIDE AMIDE GROUP).
FT MOD_RES 217 AMIDATION (G-218 PROVIDE AMIDE GROUP).
FT MOD_RES 232 AMIDATION (G-233 PROVIDE AMIDE GROUP).
FT MOD_RES 247 AMIDATION (G-248 PROVIDE AMIDE GROUP).
FT MOD_RES 262 AMIDATION (G-263 PROVIDE AMIDE GROUP).
FT MOD_RES 277 AMIDATION (G-278 PROVIDE AMIDE GROUP).
FT MOD_RES 292 AMIDATION (G-293 PROVIDE AMIDE GROUP).
FT MOD_RES 307 AMIDATION (G-308 PROVIDE AMIDE GROUP).
FT MOD_RES 322 AMIDATION (G-323 PROVIDE AMIDE GROUP).
FT MOD_RES 337 AMIDATION (G-338 PROVIDE AMIDE GROUP).
FT MOD_RES 352 AMIDATION (G-353 PROVIDE AMIDE GROUP).
FT MOD_RES 368 AMIDATION (G-369 PROVIDE AMIDE GROUP).
FT MOD_RES 384 AMIDATION (G-385 PROVIDE AMIDE GROUP).
FT MOD_RES 400 AMIDATION (G-401 PROVIDE AMIDE GROUP).
FT MOD_RES 416 AMIDATION (G-417 PROVIDE AMIDE GROUP).
FT MOD_RES 432 AMIDATION (G-433 PROVIDE AMIDE GROUP).
FT MOD_RES 448 AMIDATION (G-449 PROVIDE AMIDE GROUP).
FT MOD_RES 464 AMIDATION (G-465 PROVIDE AMIDE GROUP).
FT MOD_RES 480 AMIDATION (G-481 PROVIDE AMIDE GROUP).
FT MOD_RES 496 AMIDATION (G-497 PROVIDE AMIDE GROUP).
FT MOD_RES 512 AMIDATION (G-513 PROVIDE AMIDE GROUP).
FT MOD_RES 529 AMIDATION (G-520 PROVIDE AMIDE GROUP).
SQ SEQUENCE 552 AA; 64483 MW; ALD04999 CRC32;

Query Match 8.18; Score 86; DB 1; Length 552;

Best Local Similarity 24.08; Pred. No. 3.76e-02;

Matches 24; Conservative 25; Mismatches 45; Indels 6; Gaps 6;

Db 357 GSDE-VDKRFMRFGKSLGTD-VDKRFMRFGKSLGTD-VNKR 412
QY 38 GSFPCHIRKPGGCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPGGCHGSEPCIIHR 97
Db 413 FMRFGSLRGTD-VDKRFMRFGKSLGTD-VDKRFMRFGK 450
QY 98 GKPFSSCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFSS 137

RESULT 8
ID Z022_XENLA STANDARD; PRT; 435 AA.
AC P18745;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE OOCYTE ZINC FINGER PROTEIN XLCOP22.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89005062.
RA KOESTER M., PIELER T., POETING A., KNOBCHEL W.;
RT "The finger motif defines a multigene family represented in the
maternal mRNA of Xenopus laevis oocytes.";
RL EMBO J. 7:1735-1741(1988).

DR PIR; S00833; S00833.
DR HSSP; P08047; LSP2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 12.
DR PFAM; PF00096; zf-C2H2; 12.
KW zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 97 120 C2H2-TYPE.
FT ZN_FING 126 148 C2H2-TYPE.
FT ZN_FING 153 176 C2H2-TYPE.
FT ZN_FING 154 176 C2H2-TYPE.
FT ZN_FING 182 204 C2H2-TYPE.
FT ZN_FING 210 232 C2H2-TYPE.
FT ZN_FING 238 260 C2H2-TYPE.
FT ZN_FING 266 288 C2H2-TYPE.

FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 322 345 C2H2-TYPE.
FT ZN_FING 351 373 C2H2-TYPE.
FT ZN_FING 379 402 C2H2-TYPE.
FT ZN_FING 408 430 C2H2-TYPE.
SQ SEQUENCE 435 AA; 49416 MW; 2203DA04 CRC32;

Query Match 8.08; Score 85; DB 1; Length 435;
Best Local Similarity 31.28; Pred. No. 5.50e-02;
Matches 24; Conservative 15; Mismatches 30; Indels 8; Gaps 8;

Db 122 RKPFS-CSCGKCFSPRSLDHQTHGKPKCCQCKSKFSVSRFLDHRTHGK 180
QY 63 RGRPFSSCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFSSCHGSEPCIIHRG-K 116

Db 181 PFS-CLECGKCFSPRSLDHQTHGKPKCCQCKSKFSVSRFLDHRTHGK 196

QY 117 PFGCHGSEPCIIHRGK 133

RESULT 9
ID PSTA_DICDI STANDARD; PRT; 1046 AA.
AC P11976;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PRESTALK PROTEIN PRECURSOR.
GN ECMB.

OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Dictyostellida; Dictyostellium.

RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=AX2.

RX MEDLINE; 88015608.

RA CECCARELLI A., MCROBBIE S.J., JERMYN K.A., DUFFY K., EARLY A.,

RA WILLIAMS J.G.;

RT "Structural and functional characterization of a Dictyostellium gene

encoding a DIF inducible, prestalk-enriched mRNA sequence.";

RL Nucleic Acids Res. 15:7463-7476(1987).

RN [2]
RP SEQUENCE OF 1-17 FROM N.A.

RX MEDLINE; 87257883.

RA AYRES K., NEUMAN W., ROWEKAMP W.G., CHUNG S.;

RT "Developmental regulation of DNase I-hypersensitive sites in

Dictyostellium discoideum.";

RL Mol. Cell. Biol. 7:1823-1829(1987).

CC -I- INDUCTION: BY THE PUTATIVE STALK-SPECIFIC MORPHOGEN DIF

(DIFFERENTIATION INDUCING FACTOR).

CC -I- SIMILARITY: THE REPEATED SEQUENCES ARE HIGHLY HOMOLOGOUS TO A

REPEAT OF THE PDE3 MRNA OF DICTYOSTELLUM DISCOIDEUM.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M16345; AAA33216.1; -
DR PIR; A26838; A26838.
DR HSSP; P00592; 3P2P.
DR DICTYDB; DD03005; ECMB.
DR PFAM; PF00526; S_mold_repeat; 42.
KW Signal; Repeat.
FT SIGNAL 1 18 PROBABLE.
FT CHAIN 19 1046 PRESTALK PROTEIN.
FT DOMAIN 31 1012 41 X 24 AA TANDEM REPEATS, CYS-RICH.
FT REPEAT 31 53 X-1.
FT REPEAT 54 77 X-2.
FT REPEAT 78 101 X-3.
FT REPEAT 102 125 A-1.
FT REPEAT 126 149 A-2.

DB 170 FMRFGLSGTDD-VDKRFMRGKSLGTD-VDKRFMRGK 207
 QY 98 GKPFSSCHGSEPCIIHRGKPFSGCHGSEPCIIHRGKPF 137

RESULT 12

ID VGNMCPMV STANDARD; PRT; 1046 AA.
 AC P03599; Q84103;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GENOME POLYPROTEIN M [CONTAINS: COAT PROTEIN VP37; COAT PROTEIN VP23].
 OS Cowpea mosaic virus (CPMV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 OC Comovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 84057775.
 RA VAN WEZENBEEK P., VERVER J., HARMSSEN J., VOS P., VAN KAMMEN A.;
 RT "Primary structure and gene organization of the middle-component RNA
 of Cowpea mosaic virus.";
 RL EMBO J. 2:941-946(1983).
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF THE TWO COAT
 CC PROTEINS.
 CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-118 IS THE
 CC INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X00729; CAA25314.1; --
 CC PIR; A04210; GNWEC.
 CC HSP; P23009; IBMV.
 CC Coat protein; Polyprotein.
 CC CHAIN 460 833
 CC 834 1046
 CC SEQUENCE 1046 AA; 116217 MW; 957EB88C CRC32;

Query Match 7.8%; Score 82; DB 1; Length 1046;
 Best Local Similarity 29.4%; Pred. No. 1.69e-01;
 Matches 15; Conservative 14; Mismatches 20; Indels 2; Gaps 2;

DB 699 RYFRELHF-EVTKMSPIKATVTLLIAPGNLSDAFGYVESPHRIVQFA 748
 QY 2 QVTKMSKPIGITELGQVYIKANSKF-IGITELSSCHGSEPCIIHRGKPGF 51

RESULT 13

ID NOTCDROME STANDARD; PRT; 2703 AA.
 AC P07207; P04154;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
 GN N.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86079539.
 RA WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 product that shares homology with proteins containing EGF-like
 repeats.";

RL Cell 43:567-581(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE; 87064624.

RA KIDD S., KELLEY M.R., YOUNG M.W.;
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship
 of the encoded protein to mammalian clotting and growth factors.";
 RL Mol. Cell. Biol. 6:3094-3108(1986).
 RN [3]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE; 85099329.

RA WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 and other developmentally regulated loci in D. melanogaster.";
 RL Cell 40:55-62(1985).
 RN [4]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE; 87257846.

RA KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
 RT "Restriction of P-element insertions at the Notch locus of Drosophila
 melanogaster.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [5]
 RP REVIEW.

RA HARRIS W.A.;
 RT "Many cell types specified by Notch function.";
 RL Curr. Biol. 1:120-122(1991).
 CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS. THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.

CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.

CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M16152; AAB59220.1; --

CC EMBL; M16153; AAB59220.1; JOINED.

CC EMBL; M16149; AAB59220.1; JOINED.

CC EMBL; M16150; AAB59220.1; JOINED.

CC EMBL; M16151; AAB59220.1; JOINED.

CC EMBL; K03508; AAA28725.1; --

CC EMBL; M13689; AAA28725.1; JOINED.

CC EMBL; K03507; AAA28725.1; JOINED.

CC EMBL; M12175; AAA74496.1; --

CC EMBL; M16025; AAA28726.1; --

CC PIR; A24420; A24420.

CC PIR; A24768; A24768.

CC PIR; A05267; A05267.

CC HSP; P00740; IIXA.

CC FLYBASE; FBgn0004647; N.

CC PROSITE; PS00010; ASX-HYDROXYL; 22.

CC PROSITE; PS00022; EGF_1; 34.

CC PROSITE; PS01186; EGF_2; 28.

CC PROSITE; PS01187; EGF_CA; 22.

CC PFAM; PF00008; EGF; 36.

CC PFAM; PF00023; ank; 6.

CC PFAM; PF00066; notch; 3.

CC Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;

CC Transmembrane; Signal; Glycoprotein.

FT SIGNAL 1 44

FT POTENTIAL.


```

FT CHAIN 45 2703 NEUROGENIC LOCUS NOTCH PROTEIN.
FT DOMAIN 45 1745 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1746 1766 POTENTIAL.
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 1451 36 X EGF-TYPE REPEATS.
FT DOMAIN 58 95 EGF-LIKE 1.
FT DOMAIN 96 136 EGF-LIKE 2.
FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 865 EGF-LIKE 21.
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23.
FT DOMAIN 946 982 EGF-LIKE 24.
FT DOMAIN 984 1020 EGF-LIKE 25.
FT DOMAIN 1022 1058 EGF-LIKE 26.
FT DOMAIN 1060 1096 EGF-LIKE 27.
FT DOMAIN 1098 1134 EGF-LIKE 28.
FT DOMAIN 1136 1181 EGF-LIKE 29.
FT DOMAIN 1183 1219 EGF-LIKE 30.
FT DOMAIN 1221 1257 EGF-LIKE 31.
FT DOMAIN 1259 1295 EGF-LIKE 32.
FT DOMAIN 1297 1335 EGF-LIKE 33.
FT DOMAIN 1337 1373 EGF-LIKE 34.
FT DOMAIN 1375 1412 EGF-LIKE 35.
FT DOMAIN 1415 1451 EGF-LIKE 36.
FT DOMAIN 1475 1593 3 X LIN/NOTCH REPEATS.
FT REPEAT 1475 1513 LIN/NOTCH 1.
FT REPEAT 1514 1553 LIN/NOTCH 2.
FT REPEAT 1554 1593 LIN/NOTCH 3.
FT DOMAIN 1896 2109 6 X ANK MOTIF REPEATS.
FT DOMAIN 2338 2568 POLY-GLN (OPA-REPEAT).
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 83 BY SIMILARITY.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 100 111 BY SIMILARITY.
FT DISULFID 105 124 BY SIMILARITY.
FT DISULFID 126 135 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 181 192 BY SIMILARITY.
FT DISULFID 186 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 241 BY SIMILARITY.
FT DISULFID 243 252 BY SIMILARITY.
FT DISULFID 259 270 BY SIMILARITY.
FT DISULFID 264 279 BY SIMILARITY.
FT DISULFID 281 290 BY SIMILARITY.
FT DISULFID 297 308 BY SIMILARITY.
FT DISULFID 302 317 BY SIMILARITY.
FT DISULFID 319 328 BY SIMILARITY.
FT DISULFID 335 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 396 BY SIMILARITY.

```

```

FT DISULFID 398 407 BY SIMILARITY.
FT DISULFID 413 424 BY SIMILARITY.
FT DISULFID 418 435 BY SIMILARITY.
FT DISULFID 437 446 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 459 474 BY SIMILARITY.
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 573 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 616 BY SIMILARITY.
FT DISULFID 611 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 654 BY SIMILARITY.
FT DISULFID 648 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 686 701 BY SIMILARITY.
FT DISULFID 703 712 BY SIMILARITY.
FT DISULFID 719 730 BY SIMILARITY.
FT DISULFID 724 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 853 BY SIMILARITY.

```

... Note: remainder of annotations omitted.

Query Match 7.5%; Score 79; DB 1; Length 2703;

Best Local Similarity 22.4%; Pred. No. 5.08e-01;

Matches 19; Conservative 26; Mismatches 31; Indels 9; Gaps 8;

Db 41 LAFANLPTVRGTALVAASCTSVGCGGTCVTLQNGKTYCACDSHYVDYCE-HRNP 99

QY 28 IGITELS-SCHGSEPCIIHRGKPGGCHGSEPCIIH-RGKPFSSC--H-GSEPCIIHRGK 82

Db 100 CNSIRCONGGTCQVTFNRNGRPGISC 124

QY 83 PFGG-CHGSEPC-IIHR-GKPFSSC 104

```

RESULT 14
ID SAPE_SARPE STANDARD; PRT; 94 AA.
AC P18313;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SAPECIN PRECURSOR.
OS Sarcophaga peregrina (Flesh fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestroidea; Sarcophagidae; Sarcophaga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89034216.
RA MATSUYAMA K., NATORI S.;
RT "Molecular cloning of cDNA for sapecin and unique expression of the
RL sapecin gene during the development of Sarcophaga peregrina.";
RN J. Biol. Chem. 263:17117-17121(1988).
RP [2]
RX SEQUENCE OF 55-94.
RX MEDLINE; 89034215.

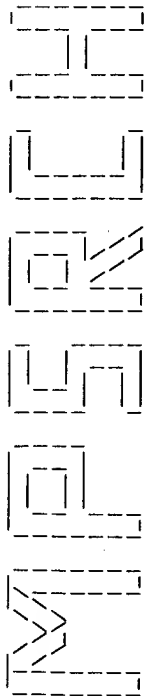
```


RA MATSUYAMA K., NATORI S.;
 RT "Purification of three antibacterial proteins from the culture medium
 of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina*.";
 RL J. Biol. Chem. 263:17112-17116(1988).
 RN [3]
 RN DISULFIDE BONDS.
 RX MEDLINE: 90292974.
 RA KUZUHARA T., NAKAJIMA Y., MATSUYAMA K., NATORI S.;
 RT "Determination of the disulfide array in sapecin, an antibacterial
 peptide of *Sarcophaga peregrina* (flesh fly).";
 RL J. Biochem. 107:514-518(1990).
 RN [4]
 RN STRUCTURE BY NMR.
 RX MEDLINE: 90382590.
 RA HANAWA H., SHIMADA I., KUZUHARA T., KOMANO H., KOHDA D., INAGAKI F.,
 NATORI S., ARATA Y.;
 RT "1H nuclear magnetic resonance study of the solution conformation of
 an antibacterial protein, sapecin.";
 RL FEBS Lett. 269:413-420(1990).
 RN [4]
 CC -1- FUNCTION: SAPECINS, WHICH ARE POTENT BACTERICIDAL PROTEINS,
 ARE PRODUCED IN RESPONSE TO INJURY. SAPECIN IS CYTOTOXIC TO
 CC GRAM-POSITIVE BACTERIA, AND IN A LESSER EXTENT AGAINST GRAM-
 CC NEGATIVE BACTERIA.
 CC -1- TISSUE-SPECIFICITY: HEMOCYTES AND FAT BODY.
 CC -1- INDUCTION: BY INJURY TO THE LARVAL CELL WALL.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD DEFENSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04053; AAA29984.1; -
 DR PIR: A31791; A31791.
 DR PIR: A31792; A31792.
 DR HSSP: P10891; LICA.
 DR PROSITE: PS00425; ARTHROPOD_DEFENSINS; 1.
 DR PFAM: PF01097; Defensin; 1.
 KW Insect immunity; Antibiotic; Cleavage on pair of basic residues;
 KW Signal.
 FT SIGNAL 1 ?
 FT PROPEP 54
 FT PEPTIDE 55 94 SAPECIN.
 FT DISULFID 57 84
 FT DISULFID 70 90
 FT DISULFID 74 92
 SQ SEQUENCE 94 AA; 9914 MW; 6BB5BD33 CRC32;
 Query Match 7.4%; Score 78; DB 1; Length 94;
 Best Local Similarity 28.3%; Pred. No. 7.27e-01;
 Matches 13; Conservative 13; Mismatches 19; Indels 1; Gaps 1;
 DB 47 ELHGRYKATCDLLSGTGNHNSAAHCLL-RGNRGYCNQKAVCV 91
 QY 15 ELGGQYIKANSFIGITELSSCHSGSEPCIIHRGKPGGCHGSEPCI 60
 RESULT 15
 ID DEFIN_PROTE STANDARD; PRT; 94 AA.
 AC P10891;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PHORMICIN PRECURSOR (INSECT DEFENSINS A AND B).
 OS *Protophormia terraenovae* (Black blowfly) (Nestling-sucking blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Cecidomyiidae; Calliphoridae; Protophormia.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 90316108.
 RA DIMARCO J.-L., ZACHARY D., HOFFMANN J.A., HOFFMANN D.,
 REICHHART J.-M.;
 RT "Insect immunity: expression of the two major inducible antibacterial
 peptides, defensin and dipterizin, in *Phormia terraenovae*.";
 RL EMBO J. 9:2507-2515(1990).
 RN [2]
 RN SEQUENCE OF 55-94.
 RX TISSUE-HEMOLYMPH;
 RX MEDLINE: 89098894.
 RA LAMBERT J., KEPPI E., DIMARCO J.-L., WICKER C., REICHHART J.-M.,
 HOFFMANN D.;
 RA DUNBAR B., LEPAGE P., VAN DORSSELAER A., HOFFMANN J.A., FOTHERGILL J.,
 HOFFMANN D.;
 RT "Insect immunity: isolation from immune blood of the dipteran *Phormia*
terraenovae of two insect antibacterial peptides with sequence
 RT homology to rabbit lung macrophage bactericidal peptides.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:262-266(1989).
 RN [3]
 RN DISULFIDE BONDS.
 RX MEDLINE: 91192047.
 RA LEPAGE P., BITSCH F., ROECKLIN D., KEPPI E., DIMARCO J.-L.,
 REICHHART J.-M., HOFFMANN J.A., ROITSCH C., VAN DORSSELAER A.;
 RT "Determination of disulfide bridges in natural and recombinant insect
 RT defensin A.";
 RL Eur. J. Biochem. 196:735-742(1991).
 RN [4]
 RN STRUCTURE BY NMR.
 RX MEDLINE: 93104264.
 RA BONMATIN J.M., GENEST M., PETIT M.-C., GINCEL E., SIMORRE J.-P.,
 CORNET B., GALLET X., CAILLE A., LABBE H., VOVELLE F., PTAK M.;
 RT "Progress in multidimensional NMR investigations of peptide and
 RT protein 3-D structures in solution. From structure to functional
 RT aspects.";
 RL Biochimie 74:825-836(1992).
 RN [5]
 RN STRUCTURE BY NMR.
 RX MEDLINE: 95393015.
 RA CORNET B., BONMATIN J.M., HETRU C., HOFFMANN J.A., PTAK M.,
 VOVELLE F.;
 RT "Refined three-dimensional solution structure of insect defensin A.";
 RL Structure 3:435-448(1995).
 CC -1- FUNCTION: RESPONSIBLE FOR THE ANTI GRAM-POSITIVE ACTIVITY OF
 CC IMMUNE HEMOLYMPH OF P.TERRAENOVAE.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD DEFENSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X55546; CAA39152.1; -
 DR PIR: A32219; A32219.
 DR PIR: B32219; B32219.
 DR PIR: S12558; S12558.
 DR PDB: LICA; 10-JUL-95.
 DR PROSITE: PS00425; ARTHROPOD_DEFENSINS; 1.
 DR PFAM: PF01097; Defensin; 1.
 KW Insect immunity; Antibiotic; Signal; 3D-structure.
 FT SIGNAL 1 23
 FT PROPEP 24 54
 FT CHAIN 55 94 PHORMICIN.
 FT DISULFID 57 84
 FT DISULFID 70 90
 FT DISULFID 74 92
 FT VARIANT 85 86 G -> R (IN DEFENSIN B).
 SQ SEQUENCE 94 AA; 10110 MW; 82F6801A CRC32;
 Query Match 7.4%; Score 78; DB 1; Length 94;
 Best Local Similarity 23.3%; Pred. No. 7.27e-01;
 Matches 14; Conservative 18; Mismatches 26; Indels 2; Gaps 2;

Db 33 HFVDGVOALKEIEPELHGRYKRATCDLLSGTGINHSACAAHCLL-RGNRGYCNKGVCV 91
QY 2 QYIKANSKFIGI-TELGQYIKANSKFIGITELSSCHGSEPCIIHRGKPFGGCHGSEPCI 60

Search completed: Sun Apr 2 17:57:52 2000
Job time : 36 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MFsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:56:05 2000; MasPar time 49.78 Seconds

Tabular output not generated.
129.811 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.pep
Perfect Score: 1057
Sequence: 1 DQYKANSKFIGITELGGQY.....FGCHGSEPCIIHRGKPFSS 137

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 34.846; Variance 64.388; scale 0.541

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	133	12.6	129	2	allergen Der f II (pF)	3.58e-09
2	133	12.6	129	2	allergen Der f II - h	3.58e-09
3	133	12.6	138	2	allergen Der f II pre	3.58e-09
4	133	12.6	138	2	allergen Der f II pre	3.58e-09
5	132	12.5	146	2	allergen Der p II pre	5.38e-09
6	104	9.8	1315	1	tentoxylisin (EC 3.4.	3.17e-04
7	91	8.6	336	2	finger protein (clone	3.54e-02
8	90	8.5	1000	2	cation efflux (AcR8/A	5.02e-02
9	89	8.4	1798	2	laminin beta-2 chain	7.11e-02
10	88	8.3	313	2	extracellular matrix	1.00e-01
11	88	8.3	379	2	larval glue protein L	1.00e-01
12	88	8.3	713	2	RVLG protein - rat	1.00e-01
13	87	8.2	597	2	FMRFamide neuropeptid	1.42e-01
14	85	8.0	435	2	finger protein (clone	2.79e-01
15	85	8.0	1046	2	prestalk protein prec	2.79e-01
16	85	8.0	1801	1	laminin beta-2 chain	2.79e-01
17	84	7.9	309	2	FMRFamide neuropeptid	3.91e-01
18	83	7.9	595	2	hypothetical protein	5.47e-01
19	83	7.9	1797	2	laminin beta-2 chain	5.47e-01
20	84	7.9	3623	2	intrinsic factor-B12	3.91e-01
21	82	7.8	1046	1	genome polyprotein M	7.62e-01
22	81	7.7	177	2	eggshell protein - fl	1.06e-00
23	81	7.7	3623	2	intrinsic factor-B12	1.06e-00

24	78	7.4	94	2	A31792	sapelin precursor - f	2.80e+00
25	78	7.4	94	2	S12558	phormicin A precursor	2.80e+00
26	78	7.4	177	2	A31204	eggshell protein - fl	2.80e+00
27	78	7.4	364	1	D43675	thymidine kinase (EC	2.80e+00
28	78	7.4	637	2	I49638	Drosophila vasa homol	2.80e+00
29	78	7.4	732	1	JC4655	acylaminoacyl-peptida	2.80e+00
30	78	7.4	1042	2	A57534	mucin (clone L31) - h	2.80e+00
31	78	7.4	1220	2	A56136	jagged protein precu	2.80e+00
32	77	7.3	199	2	E71718	hypothetical protein	3.86e+00
33	77	7.3	261	2	S77519	hypothetical protein	3.86e+00
34	77	7.3	1764	2	T15171	hypothetical protein	3.86e+00
35	77	7.3	2517	2	S58380	probable RNA-directed	3.86e+00
36	77	7.3	3566	1	A40701	tenascin-X precursor	3.86e+00
37	76	7.2	107	2	T04306	Rikria protein - rice	5.30e+00
38	76	7.2	107	2	T04305	Rikria protein - rice	5.30e+00
39	76	7.2	123	2	G48677	Ig heavy chain V-D-J	5.30e+00
40	76	7.2	154	2	T06396	isoprenylated protein	5.30e+00
41	76	7.2	755	2	A44315	cartilage oligomeric	5.30e+00
42	75	7.1	200	2	A26637	neurogenic repetitive	7.25e+00
43	74	7.0	313	2	T04776	hypothetical protein	9.89e+00
44	74	7.0	453	2	T05896	pyridoxal-phosphate-d	9.89e+00
45	74	7.0	4006	2	T09070	probable tenascin X -	9.89e+00

ALIGNMENTS

RESULT	ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	#authors	#journal	#title	#cross-references	#accession	#molecule_type	#residues	CLASSIFICATION	SUMMARY
	JU0394	allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)		30-Sep-1991	JU0394	PS0417	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haide, M.; Okudaira, H.	Agric. Biol. Chem. (1991) 55:1233-1238	Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.	#cross-references MUID:91291341	JU0394	molecule_type mRNA	1-129	#label YUU	
	JU0394	allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)		17-Mar-1999	JU0394	PS0417	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haide, M.; Okudaira, H.	Agric. Biol. Chem. (1991) 55:1233-1238	Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.	#cross-references MUID:91291341	JU0394	molecule_type mRNA	1-129	#label YUU	
	JU0394	allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)		30-Sep-1991	JU0394	PS0417	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haide, M.; Okudaira, H.	Agric. Biol. Chem. (1991) 55:1233-1238	Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.	#cross-references MUID:91291341	JU0394	molecule_type mRNA	1-129	#label YUU	
	JU0394	allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)		30-Sep-1991	JU0394	PS0417	Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haide, M.; Okudaira, H.	Agric. Biol. Chem. (1991) 55:1233-1238	Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.	#cross-references MUID:91291341	JU0394	molecule_type mRNA	1-129	#label YUU	

Query Match 12.6%; Score 133; DB 2; Length 129;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db	20	GCHGSDPCIIHRGKPFPT	36
QY	52	GCHGSEPCIIHRGKPFPS	68

RESULT	ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	#authors	#journal	#title	#cross-references	#accession	#molecule_type	#residues
	A61501	allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)		07-Oct-1994	A61501		Trudinger, M.; Chua, K.Y.; Thomas, W.R.	Clin. Exp. Allergy (1991) 21:33-37	CDNA encoding the major mite allergen Der f II.	#cross-references MUID:91215495	A61501	molecule_type mRNA	1-129
	A61501	allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)		13-Sep-1998	A61501		Trudinger, M.; Chua, K.Y.; Thomas, W.R.	Clin. Exp. Allergy (1991) 21:33-37	CDNA encoding the major mite allergen Der f II.	#cross-references MUID:91215495	A61501	molecule_type mRNA	1-129
	A61501	allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)		07-Oct-1994	A61501		Trudinger, M.; Chua, K.Y.; Thomas, W.R.	Clin. Exp. Allergy (1991) 21:33-37	CDNA encoding the major mite allergen Der f II.	#cross-references MUID:91215495	A61501	molecule_type mRNA	1-129
	A61501	allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)		13-Sep-1998	A61501		Trudinger, M.; Chua, K.Y.; Thomas, W.R.	Clin. Exp. Allergy (1991) 21:33-37	CDNA encoding the major mite allergen Der f II.	#cross-references MUID:91215495	A61501	molecule_type mRNA	1-129

```

CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476

Query Match 12.6%; Score 133; DB 2; Length 129;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSDPCIHRGKPF 36
|||||:|||||:
QY 52 GCHGSEPCIHRGKPF 68

RESULT 3
ENTRY #type fragment
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides farinae) (fragment)
DATE #formal_name Dermatophagoides farinae
12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS B61241
REFERENCE J00395
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der p II.
#cross-references MUID:92040281
#accession B61241
##molecule_type mRNA
##residues 1-138 #label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
1-9 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der p II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 12.6%; Score 133; DB 2; Length 138;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 29 GCHGSDPCIHRGKPF 45
|||||:|||||:
QY 52 GCHGSEPCIHRGKPF 68

RESULT 4
ENTRY #type fragment
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides farinae) (fragment)
DATE #formal_name Dermatophagoides farinae
12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS A61241
REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der p II.
#cross-references MUID:92040281
#accession A61241
##molecule_type mRNA
##residues 1-138 #label YUU
##note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
1-9 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der p II #status experimental #label
MAT
SUMMARY #length 138 #checksum 2894

Query Match 12.6%; Score 133; DB 2; Length 138;

```

```

Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 29 GCHGSDPCIHRGKPF 45
|||||:|||||:
QY 52 GCHGSEPCIHRGKPF 68

RESULT 5
ENTRY #type complete
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides pteronyssinus)
DATE #formal_name Dermatophagoides pteronyssinus
03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A60381
REFERENCE A60381
#authors Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;
Stewart, G.A.; Thomas, W.R.
#journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
#title Isolation of cDNA coding for the major mite allergen Der p II
by IgE plaque immunoassay.
#cross-references MUID:90256301
#accession A60381
##status not compared with conceptual translation
##molecule_type mRNA
##residues 1-146 #label CHU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-17
1-17 #domain signal sequence #status predicted #label SIG\
18-146 #product allergen Der p II #status predicted #label MAT
SUMMARY #length 146 #molecular-weight 15999 #checksum 25

Query Match 12.5%; Score 132; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 5.38e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 37 GCHGSEPCIHRGKPF 52
|||||:|||||:
QY 52 GCHGSEPCIHRGKPF 67

RESULT 6
ENTRY #type complete
TITLE tentoxylisin {EC 3.4.24.68} precursor - Clostridium tetani
ALTERNATE_NAMES tetanus neurotoxin
ORGANISM #formal_name Clostridium tetani
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
18-Jun-1999
ACCESSIONS A25689; A25757; A25194; B25194; A60759; S69348; S09364
REFERENCE A25689
#authors Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels,
J.; Weller, U.; Hudel, M.; Habermann, E.; Niemann, H.
#journal EMBO J. (1986) 5:2495-2502
#title Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
##molecule_type DNA
##residues 1-1315 #label EIS
##cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
##molecule_type DNA
##residues 1-1315 #label FAI
##cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
##experimental_source strain CN3911
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.;

```

Thomson, R.O.
J. Bacteriol. (1986) 165:21-27
Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in *Escherichia coli*.
#cross-references MUID:86085672
#accession A25194
#molecule_type DNA
#residues 743-1315 #label FA2
#cross-references GB:M12739; NID:gl44920; PIDN:AAA23282.1; PID:gl44921
#accession B25194
#molecule_type protein
#residues 865-894 #label FA3
REFERENCE A60759
#authors Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
#journal Infect. Immun. (1989) 57:3588-3593
#title Isolation, purification, and characterization of fragment B, the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 #label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell epitopes.
#cross-references MUID:89093918
#contents annotation: epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; Dasgupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
#molecule_type protein
#residues 2-31 #label DFF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B.C) chains that are covalently linked by an interchain disulfide bond (the individual chains are not toxic when separated). The amino end of the heavy chain (fragment B) can be separated from the carboxyl end (fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds to gangliosides and may target the toxin to the motor end plate. Fragment A is a zinc-dependent endopeptidase.
COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is internalized, and moves by retrograde transport up the axon into the spinal cord where it can move between postsynaptic and presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin (vesicle-associated membrane protein 2).
FUNCTION blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
CLASSIFICATION superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
FEATURE 2-457
#product tentoxylisin light chain (fragment A) #status predicted #label Ttn
#product tentoxylisin heavy chain (fragment B.C) #status predicted #label TtnH

experimental #label TtnH
: #domain channel forming (fragment B) #status predicted
#label TtnH
#domain ganglioside binding (fragment C) #status predicted
#binding_site zinc (His) #status predicted
#active_site Glu #status predicted
#length 1315 #molecular_weight 150681 #checksum 4853
Query Match 9.8%; Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 3.17e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 830 QYIKANSKFIGITEL 844
Qy 2 QYIKANSKFIGITEL 16
RESULT 7
ENTRY S06573 #type fragment
TITLE finger protein (clone XicGF48-2) - African clawed frog (fragment)
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 31-Dec-1993
ACCESSIONS S06573
REFERENCE Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeting, A.; Knoechel, W.
#journal J. Mol. Biol. (1989) 208:639-659
#title Second-order repeats in Xenopus laevis finger proteins.
#cross-references MUID:90040698
#accession S06573
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-336 #label NIE
KEYWORDS DNA binding; zinc finger
SUMMARY #length 336 #checksum 1745
Query Match 8.6%; Score 91; DB 2; Length 336;
Best Local Similarity 36.7%; Pred. No. 3.54e-02;
Matches 22; Conservative 9; Mismatches 24; Indels 5; Gaps 5;
Db 101 KFSRLTEHRKCHTEGFEVSCGKFAQSPQLDLHMRHTGKPFCS-CSDGKGFQRA 159
QY 26 KFIGITELSSCH-GSEPCI-IHRGKPGGCHGSEPCI-IHRG-KPFSSCHGSEPCIHRG 81
RESULT 8
ENTRY F70368 #type complete
TITLE cation efflux (AcrB/AcrD/AcrF family) - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
ACCESSIONS F70368
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Anjay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98196666
#accession F70368
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1000 #label AQF
#cross-references GB:AE000707; NID:g2983342; PID:g2983346; GB:AE000657
#experimental_source strain VF5
GENETICS
#gene acrD4

```

SUMMARY          #length 1000 #molecular-weight 112871 #checksum 4834

Query Match      8.5%; Score 90; DB 2; Length 1000;
Best Local Similarity 48.5%; Pred. NO. 5.02e-02;
Matches 16; Conservative 6; Mismatches 9; Indels 2; Gaps 2;

Db 739 YIKAEEDPVKEIHLNRYIIRANGELIPITEL 771
      ||||| : | : | : ||| : | : | : |||
Qy 3 YIKANSKFI-GITELGGQYIKA-NSKFIGITEL 33

RESULT 9
ENTRY   S53869      #type complete
TITLE   laminin beta-2 chain precursor (version 2) - human
ALTERNATE_NAMES s-laminin
ORGANISM #formal_name Homo sapiens #common_name man
DATE     27-Oct-1995 #sequence_revision 23-Feb-1996 #text_change
        24-Sep-1998

ACCESSIONS S53869
REFERENCE  #authors Iivanainen, A.; Vuolteenaho, R.; Sainio, K.; Eddy, R.; Shows,
            T.B.; Sariola, H.; Tryggvason, K.
            #journal Matrix Biol. (1994) 14:489-497
            #title The human laminin beta-2 chain (S-laminin): structure,
                    expression in fetal tissues and chromosomal assignment of
                    the LAMB2 gene.
            #accession S53869
            #molecule_type mRNA
            #residues 1-1798 #label IIV

GENETICS
#gene GDB:LAMB2
#cross-references GDB:132363; OMIM:150325
#map_position 3p21.3-3p21.2
CLASSIFICATION #superfamily laminin beta-1 chain; laminin-type EGF-like
                homology
KEYWORDS basement membrane; extracellular matrix; glycoprotein;
          heterotrimer

FEATURE
1-32     #domain signal sequence #status predicted #label SIG\
33-1798  #product laminin beta-2 chain #status predicted #label
          MAY\
283-344  #domain laminin-type EGF-like homology #label LE01\
347-407  #domain laminin-type EGF-like homology #label LE02\
410-467  #domain laminin-type EGF-like homology #label LE03\
470-519  #domain laminin-type EGF-like homology #label LE04\
522-552  #domain laminin-type EGF-like homology #status atypical
          #label LE05\
783-828  #domain laminin-type EGF-like homology #label LE06\
831-874  #domain laminin-type EGF-like homology #label LE07\
877-924  #domain laminin-type EGF-like homology #label LE08\
927-983  #domain laminin-type EGF-like homology #label LE09\
986-1035 #domain laminin-type EGF-like homology #label LE10\
1038-1092 #domain laminin-type EGF-like homology #label LE11\
1095-1140 #domain laminin-type EGF-like homology #label LE12\
1143-1187 #domain laminin-type EGF-like homology #label LE13\
SUMMARY   #length 1798 #molecular-weight 195993 #checksum 6417

Query Match      8.4%; Score 89; DB 2; Length 1798;
Best Local Similarity 29.9%; Pred. NO. 7.11e-02;
Matches 23; Conservative 17; Mismatches 29; Indels 8; Gaps 8;

Db 1085 NLTSGHGCPACHPSPARGPTCEFTGQCHC-RA-GFGG-RTSCCOELHWGDPGLQCH 1141
Qy 32 ELSCSGSEPCIIHRGKPFPG-GCHG-SEPCIIHRGKPFSSCHGSEPC-IIHRGKPFPGCH 88

Db 1142 ACD-CD-SRGIDTPQCH 1156
      : : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 89 GSEPCIIHRGKPFSSCH 105

RESULT 10
ENTRY   S44208      #type fragment
TITLE   extracellular matrix protein B - Dictyostellium minutum

```

```

(fragment)
#formal_name Dictyostellium minutum
07-Oct-1994 #sequence_revision 26-Jul-1996 #text_change
10-Sep-1997
ACCESSIONS S44208
REFERENCE  #authors van Es, S.; Nieuwenhuijsen, B.W.; Lenouvel, F.; van Deursen,
            E.M.; Schaap, P.
            #submission Submitted to the EMBL Data Library, April 1994
            #description Universal signals control slime mold stalk information.
            #accession S44208
            #molecule_type DNA
            #residues 1-313 #label VAN
            #cross-references EMBL:X78948; NID:g474918; PID:g474919
GENETICS
#introns 24/1
KEYWORDS extracellular matrix
SUMMARY   #length 313 #checksum 3212

Query Match      8.3%; Score 88; DB 2; Length 313;
Best Local Similarity 21.9%; Pred. No. 1.00e-01;
Matches 21; Conservative 18; Mismatches 51; Indels 6; Gaps 5;

Db 84 NCDNNACTLSDCKGCGCIHTPMNCDHNPTCLDSCDGKGCCT-H--TPMN-CDDKNAC 139
      : | : | : | | | : | : | : | : | : | : | : | : | : | : | : |
Qy 35 SCHGSEPCIIHRGKPFPGG-HGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFGGCHGSEPC 93

Db 140 TVDSCNSTGCSHTPISCDDYNSTVSDCNCNGCV 175
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 94 IHRGKPFSSC-HGSEPCIIHRGKPFPGCHGSEPC 128

RESULT 11
ENTRY   S50125      #type complete
TITLE   larval glue protein Lgp3 precursor - fruit fly (Drosophila
          virilis)
ORGANISM #formal_name Drosophila virilis
DATE     14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
        20-Aug-1999
ACCESSIONS S50125; B60095; S44059
REFERENCE  #authors Lantlo, W.; Swida, U.; Kress, H.
            #journal Biochim. Biophys. Acta (1994) 1219:576-580
            #title Molecular cloning of the Drosophila virilis larval glue
                    protein gene Lgp-3 and its comparative analysis with other
                    Drosophila glue protein genes.
            #cross-references MUID:95002181
            #accession S50125
            #molecule_type DNA
            #residues 1-379 #label LAN
            #cross-references EMBL:229565; NID:g450901; PIDN:CAA82671.1;
                    PID:g450902
REFERENCE  A60095
            #authors Swida, U.; Lucka, L.; Kress, H.
            #journal Development (1990) 108:269-280
            #title Glue protein genes in Drosophila virilis: their organization,
                    developmental control of transcription and specific mRNA
                    degradation.
            #cross-references MUID:90276249
            #accession B60095
            #molecule_type DNA
            #residues 281-379 #label SWI
            #cross-references GB:X76203; NID:g433481; PIDN:CAA53797.1; PID:e929940;
                    PID:g1335699
GENETICS
#gene FlyBase:Lgp3; Lgp-3
#map_position X16A
#introns 10/1
FUNCTION fixes the pupal case to a substrate
CLASSIFICATION #superfamily salivary glue protein
KEYWORDS glycoprotein; salivary gland; tandem repeat
FEATURE

```

[illegible]

```
##residues      1-1046 #label CEC
##experimental_source strain Ax-2

FEATURE
1-18          #domain signal sequence #status predicted #label SIG\
19-1046      #product prestalk protein #status predicted #label MAT
SUMMARY      #length 1046 #molecular-weight 108823 #checksum 3743

Query Match      8.0%; Score 85; DB 2; Length 1046;
Best Local Similarity 18.8%; Pred. No. 2.79e-01;
Matches 18; Conservative 28; Mismatches 44; Indels 6; Gaps 5;

Db 121 SCDDNNPCTVDCSNSGTGCHTPIINVVHNACTEDKC--IQLGVVTH-TPIA-CDDKNAC 176
QY 35 SCHGSEPCIHR-GKPFGGCHGSEPCIHRGKPFSSCHGSEPCIHRGKPFGGCHGSEPC 93
      ||::||::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db 177 TVDSCSNGTGCCHTPLSCDDNNPCTVDCSNSGTGV 212
QY 94 IHR-GKPFSSCHGSEPCIHRGKPFGGCHGSEPCI 128
      : :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
```

Search completed: Sun Apr 2 17:56:59 2000
Job time : 54 secs.

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:59:49 2000; MasPar time 3.75 Seconds
Tabular output not generated. 472.933 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.pep
Perfect Score: 1057
Sequence: 1 DOYIKANSKFIGITELGGQY.....FGCHGSEPCIIHRGKPFPSR 137

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 23.504; Variance 96.885; scale 0.243

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	134	12.7	61	2	US-08-478-Sequence 162, Applicat	2.05e-04
2	134	12.7	61	2	US-08-482-Sequence 162, Applicat	2.05e-04
3	134	12.7	84	2	US-08-478-Sequence 161, Applicat	2.05e-04
4	134	12.7	84	2	US-08-482-Sequence 161, Applicat	2.05e-04
5	134	12.7	92	2	US-08-478-Sequence 160, Applicat	2.05e-04
6	134	12.7	92	2	US-08-482-Sequence 160, Applicat	2.05e-04
7	134	12.7	129	2	US-08-478-Sequence 157, Applicat	2.05e-04
8	134	12.7	129	2	US-08-478-Sequence 159, Applicat	2.05e-04
9	134	12.7	129	2	US-08-482-Sequence 159, Applicat	2.05e-04
10	134	12.7	129	2	US-08-482-Sequence 157, Applicat	2.05e-04
11	134	12.7	130	2	US-08-478-Sequence 158, Applicat	2.05e-04
12	134	12.7	130	2	US-08-482-Sequence 158, Applicat	2.05e-04
13	133	12.6	129	1	US-08-462-Sequence 8, Applicatio	2.52e-04
14	133	12.6	129	3	PCT-US93-0Sequence 13, Applicati	2.52e-04
15	133	12.6	129	3	PCT-US93-0Sequence 8, Applicatio	2.52e-04
16	133	12.6	129	1	US-08-461-Sequence 13, Applicati	2.52e-04
17	133	12.6	129	1	US-08-461-Sequence 8, Applicatio	2.52e-04
18	133	12.6	129	1	US-08-461-Sequence 13, Applicati	2.52e-04
19	133	12.6	129	1	US-08-461-Sequence 8, Applicatio	2.52e-04
20	133	12.6	129	1	US-07-945-Sequence 8, Applicatio	2.52e-04
21	133	12.6	129	1	US-08-462-Sequence 13, Applicati	2.52e-04
22	133	12.6	129	2	US-08-478-Sequence 8, Applicatio	2.52e-04
23	133	12.6	129	1	US-07-945-Sequence 13, Applicati	2.52e-04

24	133	12.6	129	2	US-08-482-Sequence 8, Applicatio	2.52e-04
25	133	12.6	142	2	US-08-910-Sequence 4, Applicatio	2.52e-04
26	133	12.6	142	1	US-08-288-Sequence 2, Applicatio	2.52e-04
27	133	12.6	142	2	US-08-905-Sequence 4, Applicatio	2.52e-04
28	133	12.6	142	2	US-08-910-Sequence 2, Applicatio	2.52e-04
29	133	12.6	142	1	US-08-288-Sequence 4, Applicatio	2.52e-04
30	133	12.6	142	2	US-08-905-Sequence 2, Applicatio	2.52e-04
31	132	12.5	25	2	US-08-482-Sequence 42, Applicati	3.11e-04
32	132	12.5	25	3	PCT-US95-0Sequence 19, Applicati	3.11e-04
33	132	12.5	25	2	US-08-478-Sequence 42, Applicati	3.11e-04
34	132	12.5	129	1	US-08-462-Sequence 12, Applicati	3.11e-04
35	132	12.5	129	1	US-07-945-Sequence 12, Applicati	3.11e-04
36	132	12.5	129	3	PCT-US93-0Sequence 12, Applicati	3.11e-04
37	132	12.5	129	1	US-08-461-Sequence 12, Applicati	3.11e-04
38	132	12.5	129	1	US-08-461-Sequence 12, Applicati	3.11e-04
39	132	12.5	146	1	US-08-462-Sequence 4, Applicatio	3.11e-04
40	132	12.5	146	1	US-07-945-Sequence 4, Applicatio	3.11e-04
41	132	12.5	146	1	US-08-461-Sequence 4, Applicatio	3.11e-04
42	132	12.5	146	2	US-08-478-Sequence 4, Applicatio	3.11e-04
43	132	12.5	146	3	PCT-US93-0Sequence 4, Applicatio	3.11e-04
44	132	12.5	146	2	US-08-482-Sequence 4, Applicatio	3.11e-04
45	132	12.5	146	1	US-08-461-Sequence 4, Applicatio	3.11e-04

ALIGNMENTS

RESULT 1
ID US-08-478-572-162 STANDARD; PRT; 61 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 162, Application US/08478572
XX
CC Sequence 162, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 12.7%; Score 134; DB 2; Length 61;
Best Local Similarity 94.1%; Pred.No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSEPCIHRGKPF 36
|||||
QY 52 GCHGSEPCIHRGKPF 68
|||||

RESULT 2
ID US-08-482-142-162 STANDARD; PRT; 61 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 162, Application US/08482142
CC
CC Sequence 162, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 435
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid

CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 84 AA; 9150 MW; 34277 CN;

Query Match 12.7%; Score 134; DB 2; Length 84;

CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 12.7%; Score 134; DB 2; Length 61;
Best Local Similarity 94.1%; Pred.No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSEPCIHRGKPF 36
|||||
QY 52 GCHGSEPCIHRGKPF 68
|||||

RESULT 3
ID US-08-478-572-161 STANDARD; PRT; 84 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 161, Application US/08478572
CC
CC Sequence 161, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 84 AA; 9150 MW; 34277 CN;

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: Internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 9
ID US-08-482-142-159 STANDARD; PRT; 129 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE
Sequence 159, Application US/08482142
Sequence 159, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA

CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: Internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 10
ID US-08-482-142-157 STANDARD; PRT; 129 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE
Sequence 157, Application US/08482142
Sequence 157, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA

CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14054 MW; 84465 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 11
ID US-08-478-572-158 STANDARD; PRT: 130 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 158, Application US/08478572
XX Sequence 158, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC

CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 130 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 130 AA; 14143 MW; 85527 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 130;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 12
ID US-08-482-142-158 STANDARD; PRT: 130 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 158, Application US/08482142
XX Sequence 158, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC

CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 130 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 130 AA; 14143 MW; 85527 CN;

Query Match 12.7%; Score 134; DB 2; Length 130;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPT 36
|||||:|||||:
QY 52 GCHGSEPCIIHRGKPS 68

RESULT 13
ID US-08-462-831-8 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
DE
XX
XX
DE
XX
XX
Sequence 8, Application US/08462831
Sequence 8, Application US/08462831
Patent No. 5552142
GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/462,831
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/462,831
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC APPLICATION NUMBER: US 580,655
CC FILING DATE: 11 SEPTEMBER 1990
CC APPLICATION NUMBER: US 458,642
CC FILING DATE: 13 FEBRUARY 1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941

CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 129 AA; 14022 MW; 84331 CN;

Query Match 12.6%; Score 133; DB 1; Length 129;
Best Local Similarity 88.2%; Pred. No. 2.52e-04;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSDPCIIHRGKPT 36
|||||:|||||:
QY 52 GCHGSEPCIIHRGKPS 68

RESULT 14
ID PCT-US93-08518-13 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
DE
XX
XX
DE
XX
XX
Sequence 13, Application PC/TUS9308518
Sequence 13, Application PC/TUS9308518
GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/08518
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 11
CC OTHER INFORMATION: /label-xaa is Asn or Ser
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 52
CC OTHER INFORMATION: /label-xaa is Thr or Ile
CC FEATURE:

CC NAME/KEY: misc feature
CC LOCATION: 54
CC OTHER INFORMATION: /label-xaa is Ile or Thr
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 76
CC OTHER INFORMATION: /label-xaa is Met or Val
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 88
CC OTHER INFORMATION: /label-xaa is Ala or Ile
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 111
CC OTHER INFORMATION: /label-xaa is Val or Ile
CC
CC
SQ SEQUENCE 129 AA; 14053 MW; 88594 CN;
Query Match 12.6%; Score 133; DB 3; Length 129;
Best Local Similarity 88.2%; Pred. No. 2.52e-04;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSDPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
||||:|||||
||||:|||||
Search completed: Sun Apr 2 17:59:55 2000
Job time : 6 secs.

CC NAME/KEY: misc feature
CC LOCATION: 54
CC OTHER INFORMATION: /label-xaa is Ile or Thr
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 76
CC OTHER INFORMATION: /label-xaa is Met or Val
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 88
CC OTHER INFORMATION: /label-xaa is Ala or Ile
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 111
CC OTHER INFORMATION: /label-xaa is Val or Ile
CC
CC
SQ SEQUENCE 129 AA; 14053 MW; 88594 CN;
Query Match 12.6%; Score 133; DB 3; Length 129;
Best Local Similarity 88.2%; Pred. No. 2.52e-04;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSDPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
||||:|||||
||||:|||||

RESULT 15
ID PCT-US93-08518-8 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX
Sequence 8, Application PC/TUS9308518
Sequence 8, Application PC/TUS9308518
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
DERMATOPHAGOIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08518
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 129 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	117	42.1	343	14	Q82847	HAY SUBUNIT OF HAEMAGG	1.33e-07
2	117	42.1	344	14	Q86899	HAI-HEMAGGLUTININ [GLO	1.33e-07
3	117	42.1	359	14	Q9YJW0	HEMAGGLUTININ PRECURSO	1.33e-07
4	117	42.1	565	14	Q82559	HEMAGGLUTININ.	1.33e-07
5	114	41.0	332	14	Q56964	HAEMAGGLUTININ (FRAGME	4.68e-07
6	114	41.0	335	14	Q91205	HAEMAGGLUTININ (FRAGME	4.68e-07
7	114	41.0	342	14	Q56963	HAEMAGGLUTININ (FRAGME	4.68e-07
8	114	41.0	343	14	Q82848	HAY SUBUNIT OF HAEMAGG	4.68e-07
9	114	41.0	343	14	Q9YJ76	HEMAGGLUTININ PRECURSO	4.68e-07
10	114	41.0	343	14	Q82846	HAY SUBUNIT OF HAEMAGG	4.68e-07
11	114	41.0	343	14	Q82843	HAY SUBUNIT OF HAEMAGG	4.68e-07
12	114	41.0	343	14	Q82844	HAY SUBUNIT OF HAEMAGG	4.68e-07
13	114	41.0	346	14	Q56858	HEMAGGLUTININ HAI SUBU	4.68e-07
14	114	41.0	346	14	Q56857	HEMAGGLUTININ HAI SUBU	4.68e-07
15	114	41.0	353	14	Q12294	HEMAGGLUTININ HAI SUBU	4.68e-07
16	114	41.0	357	14	Q56856	HENAGGLUTININ HAI SUBU	4.68e-07
17	114	41.0	360	14	Q9YJY3	HENAGGLUTININ PRECURSO	4.68e-07
18	114	41.0	360	14	Q9YJY8	HENAGGLUTININ PRECURSO	4.68e-07
19	114	41.0	362	14	Q9YJY9	HENAGGLUTININ PRECURSO	4.68e-07
20	114	41.0	362	14	Q9YJY6	HENAGGLUTININ PRECURSO	4.68e-07

RESULT	2		PRT;	344 AA.
Q86899;		PRELIMINARY;		
IAD				
OC				
Q86899;				
DT				
DT				
DT				
DT				
DT				
DT				
DT				
DE				
DE				
GN				
GN				
OS				
OC				

```

DR HSP: P03437; 2VIU.
DR PFAM: PF00509; Hemagglutinin; 1.
DR PRINTS: PRO0329; HEMAGGLUTN12.
DR PRINTS: PRO0330; HEMAGGLUTN1.
DR PRINTS: PRO0331; HEMAGGLUTN2.
DR SQ SEQUENCE 565 AA; 63633 MW; A847E68B CRC32;

Query Match          42.1%; Score 117; DB 14; Length 565;
Best Local Similarity 63.3%; Pred.No. 1.33e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db    310 QNVNKVTYGK-CPKYIKQNTLKLATGMNRV 338
      || :::|||:|||||:|||||:|
QY    6 QNTLKLTAKGGPKPYKQNTLKLATGKGKV 35

RESULT 5
ID O56964 PRELIMINARY; PRT; 332 AA.
AC O56964;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-A/EQUINE/BERLIN/1/91(H3N8);
RA ILOBI C.P., NICOLSON C., TAYLOR J., MUMFORD J.A., WOOD J.M.,
RA ROBERTSON J.S.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AJ223196; CAAL1171.1; -
DR HSP: P03437; 2VIU.
DR PFAM: PF00509; Hemagglutinin; 1.
DR NON_TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >332 HA2 HAEMAGGLUTININ.
FT NON_TER 332 332
FT NON_TER
SQ SEQUENCE 332 AA; 36858 MW; C291A548 CRC32;

Query Match          41.0%; Score 114; DB 14; Length 332;
Best Local Similarity 60.0%; Pred.No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db    295 QNVNKVTYGK-CPKYIRONTLKLATGMNRV 323
      || :::|||:|||||:|||||:|
QY    6 QNTLKLTAKGGPKPYKQNTLKLATGKGKV 35

RESULT 6
ID O91205 PRELIMINARY; PRT; 335 AA.
AC O91205;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-A/EQUINE/BERLIN/4/89(H3N8);
RX MEDLINE; 98309066.
RA ILOBI C.P., NICOLSON C., TAYLOR J., MUMFORD J.A., WOOD J.M.,
RA ROBERTSON J.S.;
RL "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RL virus and comparison with laboratory derived viruses.";
RT Arch. Virol. 143:891-901(1998).
DR EMBL: AJ223195; CAAL1170.1; -
DR HSP: P03437; 2VIU.

```

Query Match 41.0%; Score 114; DB 14; Length 343;

```

QY      6 QNTLKLATGKKGPYYVQNTLKLTATGKKGV 35

RESULT 13
ID O56858 PRELIMINARY; PRT; 346 AA.
AC O56858;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HEMAGGLUTININ HAL SUBUNIT PRECURSOR (FRAGMENT).
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUI 2/BOLLNAS/96;
CA OXBURGH L., AKERBLOM L., FRIDBERGER T., KLINGEBORN B., LINNE T.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR ENBL; Y14060; CAA74388.1; -.
DR HSSP; P03437; 2VIU.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Signal; Hemagglutinin.
FT SIGNAL 1 15 POTENTIAL.
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 38449 MW; 77FE5F34 CRC32;

Query Match 41.0%; Score 114; DB 14; Length 346;
Best Local Similarity 60.0%; Pred.No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGK-CPKYIRONTLKLATGMNRY 338
|| : : || : : : : || : : : : |
| : : || : : : : || : : : : |
QY 6 QNTLKLATGKKGPYYVQNTLKLTATGKKGV 35

RESULT 14
ID O56857 PRELIMINARY; PRT; 346 AA.
AC O56857;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HEMAGGLUTININ HAL SUBUNIT PRECURSOR (FRAGMENT).
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUI 2/ALVDALEN/96;
CA OXBURGH L., AKERBLOM L., FRIDBERGER T., KLINGEBORN B., LINNE T.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR ENBL; Y14059; CAA74387.1; -.
DR HSSP; P03437; 2VIU.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Signal; Hemagglutinin.
FT SIGNAL 1 15 POTENTIAL.
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 38449 MW; 77FE5F34 CRC32;

Query Match 41.0%; Score 114; DB 14; Length 346;
Best Local Similarity 60.0%; Pred.No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGK-CPKYIRONTLKLATGMNRY 338
|| : : || : : : : || : : : : |
| : : || : : : : || : : : : |
QY 6 QNTLKLATGKKGPYYVQNTLKLTATGKKGV 35

RESULT 15
ID O12294 PRELIMINARY; PRT; 353 AA.
AC O12294;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

```

DE HEMAGGLUTININ HA1 SUBUNIT (FRAGMENT).
 OS Equine influenza virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/EQUI 2/VISINGSO /90;
 RA OXBURGH L.H.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/EQUI 2/VISINGSO /90;
 RX MEDLINE: 95159661.
 RA OXBURGH L., BERG M., KLINGEBORN B., EMMOTH E., LINNE T.;
 RT "Evolution of H3N8 equine influenza virus from 1963 to 1991."
 RL Virus Res. 34:153-165(1994).
 DR EMBL; Y14056; CAA74384.1; .
 DR HSSP; P03437; 2VIU.
 DR PFAM; PF00509; Hemagglutinin; 1.
 KW Hemagglutinin.
 FT NON_TER 353
 SQ SEQUENCE 353 AA; 39112 MW; 4A3D7FA8 CRC32;

Query Match 41.0%; Score 114; DB 14; Length 353;
 Best Local Similarity 60.0%; Pred.No. 4.68e-07;
 Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNVTYTK-CPKYIRONTLKLATGMNV 338
 |||::|||::|::|::|::|::|::|::|
 Qy 6 QNTLKLATGKKGPKYKQNTLKLATGKKG 35

Search completed: Sun Apr 2 18:04:44 2000
 Job time : 72 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:02:45 2000; MasPar time 28.23 Seconds
Tabular output not generated. 42.313 Million cell updates/sec

Title: >US-09-362-731-4
Description: (1-40) from US09362731.ppep
Perfect Score: 278
Sequence: 1 PKYVQNTLKLATGKKGPKYVQNTLKLATGKKGVIIGIK 40

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:Swissprot

Statistics: Mean 32.905; Variance 55.156; scale 0.597

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description	Pred. No.
1	117	42.1	565	1	HEMA_IAHNM HEMAGGLUTININ PRECURSO	2.99e-08
2	117	42.1	565	1	HEMA_IAHFO HEMAGGLUTININ PRECURSO	2.99e-08
3	117	42.1	565	1	HEMA_IAHSA HEMAGGLUTININ PRECURSO	2.99e-08
4	116	41.7	565	1	HEMA_IAHAL HEMAGGLUTININ PRECURSO	4.64e-08
5	115	41.4	565	1	HEMA_IAHUR HEMAGGLUTININ PRECURSO	7.19e-08
6	115	41.4	565	1	HEMA_IAHMI HEMAGGLUTININ PRECURSO	7.19e-08
7	114	41.0	565	1	HEMA_IAHSU HEMAGGLUTININ PRECURSO	1.11e-07
8	114	41.0	565	1	HEMA_IAHK7 HEMAGGLUTININ PRECURSO	1.11e-07
9	114	41.0	565	1	HEMA_IAHK6 HEMAGGLUTININ PRECURSO	1.11e-07
10	114	41.0	565	1	HEMA_IAHTE HEMAGGLUTININ PRECURSO	1.11e-07
11	114	41.0	565	1	HEMA_IAHTO HEMAGGLUTININ PRECURSO	1.11e-07
12	114	41.0	566	1	HEMA_IAUDO HEMAGGLUTININ PRECURSO	1.11e-07
13	110	39.6	328	1	HEMA_IACU7 HEMAGGLUTININ (CONTAIN	6.30e-07
14	110	39.6	328	1	HEMA_IACU6 HEMAGGLUTININ (CONTAIN	6.30e-07
15	110	39.6	550	1	HEMA_IADH6 HEMAGGLUTININ PRECURSO	6.30e-07
16	110	39.6	550	1	HEMA_IADH5 HEMAGGLUTININ PRECURSO	6.30e-07
17	110	39.6	550	1	HEMA_IADH4 HEMAGGLUTININ (CONTAIN	6.30e-07
18	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
19	110	39.6	550	1	HEMA_IADH7 HEMAGGLUTININ PRECURSO	6.30e-07
20	110	39.6	550	1	HEMA_IADH3 HEMAGGLUTININ PRECURSO	6.30e-07
21	110	39.6	550	1	HEMA_IADH1 HEMAGGLUTININ PRECURSO	6.30e-07
22	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
23	110	39.6	550	1	HEMA_IADHM HEMAGGLUTININ PRECURSO	6.30e-07

24	110	39.6	566	1	HEMA_IAME2 HEMAGGLUTININ PRECURSO	6.30e-07
25	110	39.6	566	1	HEMA_IACZO HEMAGGLUTININ PRECURSO	6.30e-07
26	110	39.6	566	1	HEMA_IATAC HEMAGGLUTININ PRECURSO	6.30e-07
27	110	39.6	566	1	HEMA_IADU3 HEMAGGLUTININ PRECURSO	6.30e-07
28	110	39.6	566	1	HEMA_IANT6 HEMAGGLUTININ PRECURSO	6.30e-07
29	110	39.6	566	1	HEMA_IANT7 HEMAGGLUTININ PRECURSO	6.30e-07
30	110	39.6	566	1	HEMA_IATV7 HEMAGGLUTININ PRECURSO	6.30e-07
31	109	39.2	567	1	HEMA_IATRO HEMAGGLUTININ PRECURSO	9.68e-07
32	108	38.8	566	1	HEMA_IATZK HEMAGGLUTININ PRECURSO	1.49e-06
33	107	38.5	550	1	HEMA_IAME6 HEMAGGLUTININ PRECURSO	2.28e-06
34	107	38.5	550	1	HEMA_IADHK HEMAGGLUTININ PRECURSO	2.28e-06
35	107	38.5	550	1	HEMA_IADH3 HEMAGGLUTININ (CONTAIN	2.28e-06
36	107	38.5	564	1	HEMA_IABUD HEMAGGLUTININ PRECURSO	2.28e-06
37	106	38.1	566	1	HEMA_IADN2 HEMAGGLUTININ PRECURSO	3.49e-06
38	105	37.8	550	1	HEMA_IAGH2 HEMAGGLUTININ PRECURSO	5.32e-06
39	105	37.8	550	1	HEMA_IADHL HEMAGGLUTININ PRECURSO	5.32e-06
40	105	37.8	566	1	HEMA_IADA3 HEMAGGLUTININ PRECURSO	5.32e-06
41	104	37.4	566	1	HEMA_IAMAO HEMAGGLUTININ PRECURSO	8.12e-06
42	103	37.1	564	1	HEMA_IASE2 HEMAGGLUTININ PRECURSO	1.24e-05
43	103	37.1	564	1	HEMA_IATRM HEMAGGLUTININ PRECURSO	1.24e-05
44	103	37.1	564	1	HEMA_IARUD HEMAGGLUTININ PRECURSO	1.24e-05
45	103	37.1	564	1	HEMA_IACKA HEMAGGLUTININ PRECURSO	1.24e-05

ALIGNMENTS

RESULT 1
ID HEMA_IAHNM STANDARD; PRT; 565 AA.
AC P16997; Q83997; Q83996;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/New Market/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOPRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use, by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M24722; AAA43107.1; ALT_SEQ.
CC FIR; E34064; HMIVES.
CC HSSP; P03437; 1HTM.
CC PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 15
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63749 MW; 0E41A00B CRC32;

```

FT CONFLICT 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63686 MW; 1727DA6B CRC32;

Query Match .. 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 2,99e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKKYTGK-CPKYIKQNTLKLATGMNRV 338
|| |:: || |::|:|||||:|:| |
QY 6 QNTLKLATGKGPKYVKQNTLKLATGKGKV 35
: :
: :

RESULT 3
ID HEMA_IAHSA STANDARD; PRT; 565 AA.
AC P16999; Q84000; Q84001;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Santiago/1/85).
OC Viruses; SRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 89204899.
RX KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement
or send an email to license@isb-sib.ch).
-----
CC EMBL; M24725; AAA43110.1; ALT_SEQ.
CC PIR; H34064; HMIVE8.
CC HSP; P03437; LHTM.
CC PFAM; PF00509; Hemagglutinin; Glycoprotein; Signal.
KW Envelope protein; Hemagglutinin;
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63665 MW; FEF4E86F CRC32;

Query Match .. 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 2,99e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKKYTGK-CPKYIKQNTLKLATGMNRV 338
|| |:: || |::|:|||||:|:| |
QY 6 QNTLKLATGKGPKYVKQNTLKLATGKGKV 35
: :
: :

RESULT 4
ID HEMA_IAHAL STANDARD; PRT; 565 AA.
AC P16994; Q83991; Q83990;
DT 01-AUG-1990 (Rel. 15, Created)

```

[illegible]

```

DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Algiers/72).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24721; AAA43100.1; ALT_SEQ.
DR PIR; D34064; HMIVP4.
DR HSSP; P03437; 1HTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 340 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 23 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63831 MW; DD483D18 CRC32;

Query Match 41.7%; Score 116; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 4.64e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKTGYGK-CPKYVKQSTLKLATGMNV 338
|||::|||:|||||:|||||:|
QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35

RESULT 5
ID HEMA_IAHUR STANDARD; PRT; 565 AA.
AC P17002; Q84007;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Uruguay/1/63).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.

```

```

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24718; AAA43114.1; ALT_SEQ.
DR PIR; A34064; HMIVEL.
DR HSSP; P03437; JHTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63604 MW; A87356D9 CRC32;

Query Match 41.4%; Score 115; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 7.19e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKTGYGK-CPKYVKQSTLKLATGMNV 338
|||::|||:|||||:|||||:|
QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35

RESULT 6
ID HEMA_IAHMI STANDARD; PRT; 565 AA.
AC P15658; Q83988; Q83987; Q67097; Q67098;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Miami/1/63).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 85133536.
RA DANIELS R.S., SKEHEL J.J., WILEY D.C.;
RT "Amino acid sequences of haemagglutinins of influenza viruses of the
RT H3 subtype isolated from horses.";
RL J. Gen. Virol. 66:457-464(1985).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```


$\frac{1}{2}$

```

Matches 18: Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMNRV 338
  || :::|| ||||:|||||||:|
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 9
ID HEMA_IAHK6 STANDARD; PRT; 565 AA.
AC P19699;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/2/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC PIR; A34065; HMIVET.
CC HSP; P03437; IHTM.
CC PFAM; PF00509; Hemagglutinin; Glycoprotein; Signal.
CC KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 POTENTIAL.
CC FT CARBOHYD 37 37 POTENTIAL.
CC FT CARBOHYD 53 53 POTENTIAL.
CC FT CARBOHYD 68 68 POTENTIAL.
CC FT CARBOHYD 78 78 POTENTIAL.
CC FT CARBOHYD 180 180 POTENTIAL.
CC FT CARBOHYD 300 300 POTENTIAL.
CC FT CARBOHYD 498 498 POTENTIAL.
CC SQ SEQUENCE 565 AA; 63610 MW; 8114DDB4 CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.1le-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMNRV 338
  || :::|| ||||:|||||||:|
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 10
ID HEMA_IAHTE STANDARD; PRT; 565 AA.
AC P17001; Q84004; Q84005;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].

```

```

GN HA.
OS Influenza A virus (strain A/Equine/Tennessee/5/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC PIR; I34064; HMIVET.
CC HSP; P03437; IHTM.
CC PFAM; PF00509; Hemagglutinin; 1.
CC KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 POTENTIAL.
CC FT CARBOHYD 37 37 POTENTIAL.
CC FT CARBOHYD 53 53 POTENTIAL.
CC FT CARBOHYD 68 68 POTENTIAL.
CC FT CARBOHYD 78 78 POTENTIAL.
CC FT CARBOHYD 180 180 POTENTIAL.
CC FT CARBOHYD 300 300 POTENTIAL.
CC FT CARBOHYD 498 498 POTENTIAL.
CC SQ SEQUENCE 565 AA; 63688 MW; CFF7B83D CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.1le-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMNRV 338
  || :::|| ||||:|||||||:|
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 11
ID HEMA_IAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M24720; AAA43111.1; ALT_SEQ.

DR PIR; C34064; HMVE3.

DR HSSP; P03437; LHTM.

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 23 23 POTENTIAL.

FT CARBOHYD 37 37 POTENTIAL.

FT CARBOHYD 53 53 POTENTIAL.

FT CARBOHYD 68 68 POTENTIAL.

FT CARBOHYD 78 78 POTENTIAL.

FT CARBOHYD 180 180 POTENTIAL.

FT CARBOHYD 300 300 POTENTIAL.

FT CARBOHYD 498 498 POTENTIAL.

SQ SEQUENCE 565 AA; 63580 MW; 9A54B199 CRC32;

Query Match

Best Local Similarity 41.0%; Score 114; DB 1; Length 565;

Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKITYGK-CPKYVKQNTLKLATGMNRI 338

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35

RESULT 12

ID HEMA_IAUDO STANDARD; PRT; 566 AA.

AC P19106; Q67022; Q67023;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;

DE HEMAGGLUTININ HA2 CHAIN].

GN HA.

OS Influenza A virus (strain A/Udorn/307/72).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group.

CC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE; 85050743.

RA YUFEROV V., KARGINOV V., SAMOKHYALOV E., CHIZHIKOV V., VASILONKO S.,

RA URYVAEV L., ZHDANOV V.M.;

RA A/Udorn/307/72 (H3N2).";

RT "Nucleotide sequence of the hemagglutinin gene of the Influenza virus

DoKl. Akad. Nauk SSSR 278:738-742(1984).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS

(HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M54895; AAA43099.1; ..

DR EMBL; M54895; CAB25932.1; ALT_TERM.

DR EMBL; M54895; CAB25933.1; ALT_INIT.

DR HSSP; P03437; LHTM.

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 24 24 POTENTIAL.

FT CARBOHYD 38 38 POTENTIAL.

FT CARBOHYD 54 54 POTENTIAL.

FT CARBOHYD 97 97 POTENTIAL.

FT CARBOHYD 181 181 POTENTIAL.

FT CARBOHYD 301 301 POTENTIAL.

FT CARBOHYD 499 499 POTENTIAL.

SQ SEQUENCE 566 AA; 63190 MW; 17EABF51 CRC32;

Query Match

Best Local Similarity 63.3%; Score 114; DB 1; Length 566;

Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 311 QNVNKITYGA-CPKYVKQNTLKLATGMNRY 339

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35

RESULT 13

ID HEMA_IAQ07 STANDARD; PRT; 328 AA.

AC P04663;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HEMAGGLUTININ [CONTAINS: HEMAGGLUTININ HA1 CHAIN] (FRAGMENT).

GN HA.

OS Influenza A virus (strain A/Qu/7/70).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group.

RN [1]

RP SEQUENCE FROM N.A.

RA SLEIGH M.J., BOTH G.W., UNDERWOOD P.A., BENDER V.J.;

RT "Antigenic drift in the hemagglutinin of the Hong Kong Influenza

subtypes: correlation of amino acid changes with alterations in viral

antigenicity.";

RL J. Virol. 37:845-853(1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS

(HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; K03338; AAA43195.1; ..

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT CHAIN 1 328

FT CARBOHYD 8 8 HEMAGGLUTININ HA1 CHAIN.

FT CARBOHYD 22 22 POTENTIAL.

FT CARBOHYD 38 38 POTENTIAL.

FT CARBOHYD 81 81 POTENTIAL.

FT CARBOHYD 165 165 POTENTIAL.

FT CARBOHYD 285 285 POTENTIAL.

FT NON_TER 328 328

SQ SEQUENCE 328 AA; 36059 MW; E6980E2B CRC32;

Query Match

Best Local Similarity 39.6%; Score 110; DB 1; Length 328;

Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 295 QNVNKITYGA-CPKYVKQNTLKLATGMNRY 323

(TM)

Result	Query	Score	Match	Length	DB	ID	Description	Pred. No.
1	117	42.1	565	1	HMIV6	hemagglutinin precurs	5.59e-07	
2	117	42.1	565	1	HMIV8	hemagglutinin precurs	5.59e-07	
3	117	42.1	565	1	HMIV5	hemagglutinin precurs	5.59e-07	
4	116	41.7	565	1	HMIV4	hemagglutinin precurs	8.30e-07	
5	115	41.4	565	1	HMIV1	hemagglutinin precurs	1.23e-06	
6	115	41.4	565	1	HMIV2	hemagglutinin precurs	1.23e-06	
7	114	41.0	565	2	S33703	hemagglutinin - influ	1.82e-06	
8	114	41.0	565	1	HMIV7	hemagglutinin precurs	1.82e-06	
9	114	41.0	565	1	HMIV3	hemagglutinin precurs	1.82e-06	
10	114	41.0	565	1	HMIV3	hemagglutinin precurs	1.82e-06	
11	114	41.0	565	1	HMIV9	hemagglutinin precurs	1.82e-06	
12	110	39.6	550	2	J01156	hemagglutinin precurs	8.64e-06	
13	110	39.6	550	2	HMIV8	hemagglutinin precurs	8.64e-06	
14	110	39.6	550	1	HMIV0	hemagglutinin precurs	8.64e-06	
15	110	39.6	550	1	HMIV5	hemagglutinin precurs	8.64e-06	
16	110	39.6	550	1	HMIV3	hemagglutinin precurs	8.64e-06	
17	110	39.6	550	2	A92979	hemagglutinin precurs	8.64e-06	
18	110	39.6	550	1	HMIV33	hemagglutinin precurs	8.64e-06	
19	110	39.6	550	1	HMIV77	hemagglutinin precurs	8.64e-06	
20	110	39.6	550	1	HMIV89	hemagglutinin precurs	8.64e-06	
21	110	39.6	566	1	HMIVD	hemagglutinin precurs	8.64e-06	
22	110	39.6	566	1	HMIVHA	hemagglutinin precurs	8.64e-06	
23	110	39.6	566	1	HMIVHM	hemagglutinin precurs	8.64e-06	

```

CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
          thiolester bond
FEATURE
1-16 : #domain signal sequence #status predicted #label SIG\
17-344 : #product hemagglutinin HA1 #status predicted #label HA1\
345-565 : #product hemagglutinin HA2 #status predicted #label HA2\
535-551 : #domain transmembrane #status predicted #label TM1\
23,37,53,78,180, :
300,498 : #binding_site carbohydrate (Asn) (covalent) #status
          predicted\
29-481,67-292, :
79-91,154-488, :
296-320 : #disulfide_bonds #status predicted\
554,561,564 : #binding_site palmitate (Cys) (covalent) #status
          predicted
SUMMARY #length 565 #molecular_weight 63749 #checksum 4207
Query Match 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 5.59e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKTGYGK-CPKYIKQNTLKLATGMNV 338
II I::II IIIIIIIIIIIIIIIII::I
QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKRGV 35

RESULT 4
ENTRY HMVB4 #type complete
TITLE hemagglutinin precursor - Influenza A virus (strain
A/equine/Algers/72[H3N8])
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name Influenza A virus
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
16-Jul-1999
ACCESSION D34064
REFERENCE A34064
#authors Kawaoka, Y.; Bean, W.J.; Webster, R.G.
#journal Virology (1989) 169:283-292
#title Evolution of the hemagglutinin of equine H3 influenza
viruses.
#cross-references MUID:89204899
#accession D34064
#molecule_type genomic RNA
#residues 1-565 #label KAW
##cross-references GB:M4721; GB:J04336; NID:g323996; PIDN:AAA43100.1;
PID:g323997
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
          thiolester bond
FEATURE
1-16 : #domain signal sequence #status predicted #label SIG\
17-344 : #product hemagglutinin HA1 #status predicted #label HA1\
345-565 : #product hemagglutinin HA2 #status predicted #label HA2\
535-551 : #domain transmembrane #status predicted #label TM1\
23,37,53,78,180, :
300,498 : #binding_site carbohydrate (Asn) (covalent) #status
          predicted\
29-481,67-292, :
79-91,154-488, :
296-320 : #disulfide_bonds #status predicted\
554,561,564 : #binding_site palmitate (Cys) (covalent) #status
          predicted
SUMMARY #length 565 #molecular_weight 63831 #checksum 4800
Query Match 41.7%; Score 116; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 8.30e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKTGYGK-CPKYIKQNTLKLATGMNV 338
II I::II IIIIIIIIIIIIIIIII::I

```

GENETICS

```

554,561,564      #binding_site palmitate (Cys) (covalent) #status
                  predicted
SUMMARY          #length 565 #molecular-weight 63702 #checksum 5766

Query Match      41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.82e-06;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGGK-CPKYVIRONTLKLATGMENV 338
|||::||| |||::|||||::| |
Qy   6 QNTLKATGKKGPVKYVQNTLKLATGKKGV 35

RESULT 10
ENTRY HMIWE9 #type complete
TITLE hemagglutinin precursor - Influenza A virus (strain
A/equine/Tokyo/71[H3N8])
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name Influenza A virus
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
16-Jul-1999
ACCESSIONS C34064
REFERENCE A34064
#authors Kawaoka, Y.; Bean, W.J.; Webster, R.G.
#journal Virology (1989) 169:283-292
#title Evolution of the hemagglutinin of equine H3 influenza
viruses.
#crosso-references MDID:89204899
#accession C34064
#molecule_type genomic RNA
#residues 1-565 #label KAW
#cross-references GB:M24720; GB:J04336; NID:g324018; PIDN:AAA43111.1;
PID:g324019

GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
cholesterol bond

FEATURE
1-16 #domain signal sequence #status predicted #label SIG\
17-344 #product hemagglutinin HA1 #status predicted #label HA1\
345-565 #product hemagglutinin HA2 #status predicted #label HA2\
535-551 #domain transmembrane #status predicted #label TM1\
23,37,53,68,78,180,
300,498 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
29-481,67-292, #disulfide_bonds #status predicted\
79-91,154-488, #binding_site palmitate (Cys) (covalent) #status
296-320 predicted
554,561,564

SUMMARY          #length 565 #molecular-weight 63580 #checksum 3980

Query Match      41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.82e-06;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKITYGGK-CPKYVQSTLKLATGMENI 338
|||::||| |||::|||||::| |
Qy   6 QNTLKATGKKGPVKYVQNTLKLATGKKGV 35

RESULT 11
ENTRY HMIWE9 #type complete
TITLE hemagglutinin precursor - Influenza A virus (strain
A/equine/Tennessee/5/85[H3N8])
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name Influenza A virus
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
16-Jul-1999
ACCESSIONS I34064
REFERENCE A34064
#authors Kawaoka, Y.; Bean, W.J.; Webster, R.G.
```


Best Local Similarity 63.3%; Pred. No. 8.64e-06; Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 295 QNVNKITYGA-CPKYVKQNTLKLATGMRNV 323
|||::| | |||||||||::|
QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKKV 35

RESULT 13
ENTRY HMIV80 #type fragment
TITLE hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name influenza A virus
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
ACCESSION B27813
REFERENCE A94363
#authors Kida, H.; Kawakita, Y.; Naeve, C.W.; Webster, R.G.
#journal Virology (1987) 159:109-119
#title Antigenic and genetic conservation of H3 influenza virus in wild ducks.
#cross-references MUID:87265458
#accession B27813
#molecule_type genomic RNA
##residues 1-550 ##label KID
##cross-references GB:M16738; NID:g324083
##note the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1 in having I37-Asn rather than Lys for codon AAG

GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

FEATURE
1-328
330-550
520-536
8,22,38,165,285,
483
14-466,52-277,
64-76,139-473,
281-305
539,546,549
#disulfide_bonds #status predicted\
#binding_site palmitate (Cys) (covalent) #status Predicted
#length 550 #checksum 4376

SUMMARY
Query Match 39.6%; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 295 QNVNKITYGA-CPKYVKQNTLKLATGMRNV 323
|||::| | |||||||||::|
QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKKV 35

RESULT 14
ENTRY HMIV15 #type fragment
TITLE hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name influenza A virus
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
ACCESSION G27813
REFERENCE A94363
#authors Kida, H.; Kawakita, Y.; Naeve, C.W.; Webster, R.G.
#journal Virology (1987) 159:109-119
#title Antigenic and genetic conservation of H3 influenza virus in wild ducks.
#cross-references MUID:87265458

```
#accession G27813
#molecule_type genomic RNA
##residues 1-550 ##label KID
##cross-references GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
thiolester bond
FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1\
330-550 #product hemagglutinin HA2 #status predicted #label HA2\
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\
14-466,52-277,
64-76,139-473,
281-305 #disulfide_bonds #status predicted\
539,546,549 #binding_site palmitate (Cys) (covalent) #status
predicted
SUMMARY #length 550 #checksum 5685
Query Match 39.6%; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 295 QNVNKITYGA-CPKYVKONTLKLATGMNV 323
QY 6 ONTLKLATGKKGPKYVKONTLKLATGKGV 35
Search completed: Sun Apr 2 18:02:28 2000
Job time : 48 secs.
#accession B29971
#molecule_type genomic RNA
##residues 1-550 ##label KID
##cross-references GB:M19057; NID:g324210
##note the sequence in GenBank entry FLAHAPB, release 106,
(PID:g324211) differs from the sequence in Fig. 1 in
omitting the nucleotides GCATTT before the final
dinucleotide
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
thiolester bond
FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1\
330-550 #product hemagglutinin HA2 #status predicted #label HA2\
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\
14-466,52-277,
64-76,139-473,
281-305 #disulfide_bonds #status predicted\
539,546,549 #binding_site palmitate (Cys) (covalent) #status
predicted
SUMMARY #length 550 #checksum 3973
```

```
Query Match 39.6%; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 295 QNVNKITYGA-CPKYVKONTLKLATGMNV 323
QY 6 ONTLKLATGKKGPKYVKONTLKLATGKGV 35
Search completed: Sun Apr 2 18:02:28 2000
Job time : 48 secs.
```

M I S R L A
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:08:41 2000; MasPar time 67.05 Seconds
33.089 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QYIKANSKFIGHTELGCGHSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-plant 10:sp-rodent 11:sp-rodent 12:sp-unclassified
13:sp-virus 14:sp-virus

Statistics: Mean 30.267; Variance 37.728; scale 0.802

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	98	40.7	145	5	095430	GROUP 2 ALLERGEN EUR M	4.91e-08
2	75	31.1	533	6	09XT71	RETINAL PIGMENT EPITHE	1.14e-02
3	75	31.1	533	6	097623	RETINAL PIGMENT EPITHE	1.14e-02
4	75	31.1	533	4	Q15518	RETINAL PIGMENT EPITHE	1.14e-02
5	75	31.1	533	6	Q05661	RETINAL PIGMENT EPITHE	1.14e-02
6	75	31.1	533	6	Q28175	RETINAL PIGMENT EPITHE	1.14e-02
7	75	31.1	533	11	070276	RETINAL PIGMENT EPITHE	1.14e-02
8	72	29.9	173	10	048704	T9J22.4 PROTEIN.	5.00e-02
9	70	29.0	116	10	09XIA6	PUTATIVE RIBOSOMAL PRO	1.31e-01
10	69	28.6	102	10	024118	40S RIBOSOMAL PROTEIN	2.12e-01
11	69	28.6	133	10	083208	F6P22.28 PROTEIN.	2.12e-01
12	69	28.6	502	10	083752	HYPOTHETICAL 55.2 KD P	2.12e-01
13	68	28.2	62	2	09X5H5	UREASE-ENHANCING FACTO	3.40e-01
14	68	28.2	382	2	032874	HOMOSERINE O-ACETYLTRA	3.40e-01
15	68	28.2	1036	1	029316	HYPOTHETICAL 112.4 KD	3.40e-01
16	68	28.2	1067	13	09YHU6	PHOSPHOLIPASE C-GAMMA-	3.40e-01
17	66	27.4	379	2	053391	HOMOSERINE O-ACETYLTRA	8.61e-01
18	65	27.0	131	7	046869	MHC CLASS II BETA CHAI	1.36e+00
19	65	27.0	356	5	025255	PERITROPIN-44 PRECURS	1.36e+00
20	65	27.0	1215	5	077202	MYOSIN-IA.	1.36e+00

21	64	26.6	38	2	O25251	HYPOTHETICAL 4.0 KD PR	2.14e+00
22	64	26.6	159	1	O9YE65	159AA LONG HYPOTHETICA	2.14e+00
23	64	26.6	583	11	O54714	PIAS3.	2.14e+00
24	63	26.1	138	10	O9XG80	PUTATIVE PHOSPHOLIPASE	3.35e+00
25	63	26.1	392	13	O9W641	ALPHA(1,3)FUCOSYLTRANS	3.35e+00
26	62	25.7	131	7	O48666	MHC CLASS II BETA CHAI	5.21e+00
27	62	25.7	149	11	O920J0	EPIDIDYMAL SECRETORY P	5.21e+00
28	62	25.7	163	2	O9X126	CONSERVED HYPOTHETICAL	5.21e+00
29	62	25.7	219	1	O28237	URIDYLATE KINASE (PYRH	5.21e+00
30	62	25.7	445	5	O45870	T27F6.6 PROTEIN.	5.21e+00
31	62	25.7	598	2	O34469	YEEB PROTEIN.	5.21e+00
32	62	25.7	565	10	O49236	VIVIPAROUS 1 HOMOLOGUE	5.21e+00
33	62	25.7	899	14	O9YTK4	ORF 63.	5.21e+00
34	61	25.3	131	14	O9YPT2	V2 PROTEIN.	8.06e+00
35	61	25.3	156	1	P95911	ORF C01013.	8.06e+00
36	61	25.3	157	5	O01698	F33E2.4 PROTEIN.	8.06e+00
37	61	25.3	208	3	O9Y7P0	RIBOFLAVIN SYNTHASE AL	8.06e+00
38	61	25.3	264	2	O9X7C6	PUTATIVE MEMBRANE PROT	8.06e+00
39	61	25.3	349	3	P78872	FISSION YEAST (FRAGMEN	8.06e+00
40	61	25.3	361	10	P93257	CINNAMYL-ALCOHOL DEHYD	8.06e+00
41	61	25.3	376	2	O9WVA3	HYPOTHETICAL 41.5 KD P	8.06e+00
42	61	25.3	433	2	O97748	FLAGELLUM-SPECIFIC ATP	8.06e+00
43	61	25.3	468	5	O9XV51	F29C12.5 PROTEIN.	8.06e+00
44	61	25.3	667	6	O46480	ENDOOLIGOPEPTIDASE A R	8.06e+00
45	61	25.3	1388	5	Q18566	SIMILAR TO DNA-DIRECTE	8.06e+00

ALIGNMENTS

RESULT 1
ID O96430 PRELIMINARY; PRT; 145 AA.
AC O96430;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Eukaryotus maynei (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcotiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Euroglyphus.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH W., HART B.J., THOMAS W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047613; AAC82349.1;
DR HSP; P49278.1A9V.
SQ SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;
Query Match 40.7%; Score 98; DB 5; Length 145;
Best Local Similarity 75.0%; Pred. No. 4.91e-08;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 36 GCKGSEPCVHRGTF 51
QY 17 GCKGSEPCNIHRGKPF 32
RESULT 2
ID O9X771 PRELIMINARY; PRT; 533 AA.
AC O9X771;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE RETINAL PIGMENT EPITHELIUM-SPECIFIC PROTEIN RPE65.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
OC Chlorocebus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA MA J.-X., ZHANG D., BROWNLEE N., RE G.R., HAZEN-MARTIN D.J.,

RA REDMOND T.M., CROUCH R.K.;
RT "Expression of RP65 mRNA in Transformed Kidney Cells and Renal Tumor Cells.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF093455; AAD42042.1; -;
SQ SEQUENCE 533 AA; 60975 MW; 0334B7D5 CRC32;
Query Match 31.1%; Score 75; DB 6; Length 533;
Best Local Similarity 44.0%; Pred. No. 1.14e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
QY 2 YIKANS-KFIGITELGGCHGSEPCN 25
RESULT 3
ID Q97623 PRELIMINARY; PRT; 533 AA.
AC Q97623;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE RETINAL PIGMENT EPITHELIUM-SPECIFIC PROTEIN RPE65.
GN RPE65.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RX MEDLINE; 99030756.
RA AGUIRRE G.D., BALDWIN V., PEARCE-KELLING S., NARFSTROM K., RAY K.,
RA ACLAND G.M.;
RT "Congenital stationary night blindness in the dog: common mutation in
RT the RPE65 gene indicates founder effect.";
RL Mol. Vis. 4:23-23(1998).
DR EMBL; AF084537; AAC72356.1; -;
SQ SEQUENCE 533 AA; 60880 MW; 58A03ABB CRC32;
Query Match 31.1%; Score 75; DB 6; Length 533;
Best Local Similarity 44.0%; Pred. No. 1.14e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
QY 2 YIKANS-KFIGITELGGCHGSEPCN 25
RESULT 4
ID Q16518 PRELIMINARY; PRT; 533 AA.
AC Q16518;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE RETINAL PIGMENT EPITHELIUM-SPECIFIC 61 KDA PROTEIN.
GN RPE65.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RX MEDLINE; 95359969.
RA NICOLETTI A., WONG D.J., KAWASE K., GIBSON L.H., YANG-PENG T.L.,
RA RICHARDS J.E., THOMPSON D.A.;
RT "Molecular characterization of the human gene encoding an abundant 61
RT kDa protein specific to the retinal pigment epithelium.";
RL Hum. Mol. Genet. 4:641-649(1995).
RN [2]
RN SEQUENCE FROM N.A.
RP MORIMURA H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U18991; AAA99012.1; -;
DR EMBL; U20510; AAC14586.1; -;

DR EMBL; U20476; AAC14586.1; JOINED.
DR EMBL; U20477; AAC14586.1; JOINED.
DR EMBL; U20478; AAC14586.1; JOINED.
DR EMBL; U20479; AAC14586.1; JOINED.
DR EMBL; U20481; AAC14586.1; JOINED.
DR EMBL; U20482; AAC14586.1; JOINED.
DR EMBL; U20484; AAC14586.1; JOINED.
DR EMBL; U20485; AAC14586.1; JOINED.
DR EMBL; U20486; AAC14586.1; JOINED.
DR EMBL; AF039868; AAC39660.1; JOINED.
DR EMBL; AF039855; AAC39660.1; JOINED.
DR EMBL; AF039856; AAC39660.1; JOINED.
DR EMBL; AF039857; AAC39660.1; JOINED.
DR EMBL; AF039858; AAC39660.1; JOINED.
DR EMBL; AF039859; AAC39660.1; JOINED.
DR EMBL; AF039860; AAC39660.1; JOINED.
DR EMBL; AF039861; AAC39660.1; JOINED.
DR EMBL; AF039862; AAC39660.1; JOINED.
DR EMBL; AF039863; AAC39660.1; JOINED.
DR EMBL; AF039864; AAC39660.1; JOINED.
DR EMBL; AF039865; AAC39660.1; JOINED.
DR EMBL; AF039866; AAC39660.1; JOINED.
DR EMBL; AF039867; AAC39660.1; JOINED.
SQ SEQUENCE 533 AA; 60947 MW; 0B8A41C3 CRC32;
Query Match 31.1%; Score 75; DB 4; Length 533;
Best Local Similarity 44.0%; Pred. No. 1.14e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
QY 2 YIKANS-KFIGITELGGCHGSEPCN 25
RESULT 5
ID Q05661 PRELIMINARY; PRT; 533 AA.
AC Q05661;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE RETINAL PIGMENT EPITHELIUM MEMBRANE RECEPTOR P63.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 365-382 AND 407-423.
RX STRAIN-STEINHOLTZ; TISSUE-EYE;
RX MEDLINE; 93388633.
RA BAVIK C.O., HELLMAN U., WERNSTEDT C., ERIKSSON U.;
RT "The retinal pigment epithelial membrane receptor for plasma retinol-
RT binding protein. Isolation and cDNA cloning of the 63-kDa protein.";
RL J. Biol. Chem. 268:20340-20346(1993).
CC -1- FUNCTION: THIS RECEPTOR BINDS PLASMA RETINOL BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN RETINAL PIGMENT
CC EPITHELIUM.
DR EMBL; X66277; CAA46988.1; -;
KW Receptor; Retinol-binding; Plasma; Membrane.
FT MOD_RES 71 BLOCKED.
SQ SEQUENCE 533 AA; 60970 MW; F77E22D7 CRC32;
Query Match 31.1%; Score 75; DB 6; Length 533;
Best Local Similarity 44.0%; Pred. No. 1.14e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
QY 2 YIKANS-KFIGITELGGCHGSEPCN 25
RESULT 6
ID Q28175 PRELIMINARY; PRT; 533 AA.

```

RESULT      8
ID O48704 PRELIMINARY; PRT; 173 AA.
AC O48704;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DDT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DDT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)
DE T9J22.4 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC eumolliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsi.
OC [1]
RRN SEQUENCE FROM N.A.
RRP STRAIN=CV. COLUMBIA;
RRC ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RRA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RRL SOMERVILLE C.N., VENTER J.C.;
RRL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RRL EMBL; AC002505; AAC14488.1; -.
DSQ SEQUENCE 173 AA; 19015 MW; 01243A3B CRC32;
Query Match 29.9%; Score 72; DB 10; Length 173;
Best Local Similarity 35.5%; Pred. No. 5,00e-02;
Matches 11; Conservative 6; Mismatches 13; Indels 1; Gaps
Ddb 88 YENVNSTINY-DLSVMGDEPCSIPEGEKF 117
| : | | : | : | : ||||| | |
QY 2 YIKANSKFIGITELGGCHSGEPCNIHRGKPF 32
          PRELIMINARY; PRT; 116 AA.
RESULT      9
ID Q9XIA6 PRELIMINARY; PRT; 116 AA.
AC Q9XIA6;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DDT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DDT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE PUTATIVE RIBOSOMAL PROTEIN.
DSQ F13F21.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsi.
OC [1]
RRN SEQUENCE FROM N.A.
RRP FEDERSPIEL N.A., PALM C.J., CONWAY A.B., CONN L., HANSEN N.F.,
RRA ALTAFI H., ARAUJO R., HUIZAR L., ROWLEY D., BUEHLER E., DUNN P.,
RRC GONZALEZ A., KREMENTETSKAIA I., KIM C., LENZ C., LI J., LIU S.,
RRA LURON S., SCHWARTZ J., SHINN P., TORIUMI M., VYSOTSKAIA V.S.,
RRL WALKER M., YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RRL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RRL EMBL; AC007504; AAD43160.1; -.
DSQ Ribosomal protein.
SEQUENCE 116 AA; 12560 MW; 08946EDD CRC32;
Query Match 29.0%; Score 70; DB 10; Length 116;
Best Local Similarity 21.9%; Pred. No. 1.31e-01;
Matches 7; Conservative 14; Mismatches 11; Indels 0; Gaps
Ddb 32 RYVKTSFKMAHDDKDAGCNIGRVKLDPSPRL 63
| | | | | : | : | : | : | : | : |
QY 1 QYIKANSFIGITELGGCHSGEPCNIHRGKPF 32
          PRELIMINARY; PRT; 102 AA.
RESULT      10
ID O24118 PRELIMINARY; PRT; 102 AA.
AC O24118;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DDT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DDT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE 40S RIBOSOMAL PROTEIN S17.

```

PETERS S.A., VAN STAVEREN M., DIRKSE W., STIEKEMA W., BANCROFT I., MEWES H.W., MAYER K.F.X., SCHUELLER C., BEVAN M.; Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

EU ARABIDOPSIS SEQUENCING PROJECT;

Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

EMBL; AL031018; CAA19819.1; -

PFAM; PF01554; UPF0013; 1.

Hypothetical protein.

SEQUENCE 502 AA; 55232 MW; B6DF0F26 CRC32;

Query Match 28.6%; Score 69; DB 10; Length 502;

Best Local Similarity 42.9%; Pred. No. 2,12e-01;

Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 382 IIGLCELGNCPQTTLGGVLRG 402

YQ :||| | | : | : ||

9 FIGITELGGCHGSEPCNIHRG 29

RESULT 13

ID Q9X5H5 PRELIMINARY; PRT; 62 AA.

Q9X5H5;

AC

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)

DE UREASE-ENHANCING FACTOR LPP.

GN LPP.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

[1]

SEQUENCE FROM N.A.

STRAIN-ATCC 43504;

RC

RX MEDLINE; 99214098.

RA MCGEE D.J., MAY C.A., GARNER R.M., HIMPSE J.M., MOBLEY H.L.T.;

RT "Isolation of Helicobacter pylori genes that modulate urease activity.";

RL J. Bacteriol. 181:2477-2484(1999).

DR EMBL; AF125214; AAD27692.1; -

SEQUENCE 62 AA; 6812 MW; FOBA2DE5 CRC32;

Query Match 28.2%; Score 68; DB 2; Length 62;

Best Local Similarity 58.8%; Pred. No. 3,40e-01;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 6 NSMFIGASLLGGCASVE 22

YQ :||| : ||| : |

6 NSKFIGITELGGCHGSE 22

RESULT 14

ID Q32874 PRELIMINARY; PRT; 382 AA.

Q32874;

AC

DT 01-JAN-1998 (TReMBLrel. 05, Created)

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)

DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)

DE HOMOSERINE O-ACETYLTRANSFERASE.

GN MLCB1779.11.

OS Mycobacterium leprae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

[1]

SEQUENCE FROM N.A.

RA HAMLIN N., CHURCHER C.M.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DT

RP

SEQUENCE FROM N.A.

RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE FROM N.A.

RP

```

RX MEDLINE: 93189700.
RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL: Z98271; CAB10992.1; -.
DR PFAM: PF00561; abhydrolase; 1.
KW Transferase.
SQ SEQUENCE 382 AA; 40298 MW; B5B9C970 CRC32;

Query Match      28.2%; Score 68; DB 2; Length 382;
Best Local Similarity 50.0%; Pred. No. 3.40e-01;
Matches 12; Conservative 4; Mismatches 6; Indels 2; Gaps 2;

Db 98 IATNVLGCGRGSTGPSLARDGKP 121
   |:::|::| |::|::|
QY 10 IGITELGGCHGS-EPCNIHR-GRP 31

RESULT 15
ID O29316 PRELIMINARY; PRT; 1036 AA.
AC O29316;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)
DE HYPOTHETICAL 112.4 KD PROTEIN.
GN AF0946.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA RICHARDSON D.L., KERLAVAGE A.R., GWINN M., HICKEY E.K., PETERSON J.D.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOWN P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL: AF001039; AAB90312.1; -.
DR TIGR: AF0946; -.
KW Hypothetical protein.
SQ SEQUENCE 1036 AA; 112425 MW; FOFDA8F4 CRC32;

Query Match      28.2%; Score 68; DB 1; Length 1036;
Best Local Similarity 30.0%; Pred. No. 3.40e-01;
Matches 9; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 543 INSEVFK-IDDLAFCSGDPDCDV-REKTY 570
   |:::| |::|::| |::|::|
QY 3 IKANSKFIGITELGGCHGSEPCNIHRGKPF 32

Search completed: Sun Apr 2 18:09:54 2000
Job time : 73 secs.

```

MUSEARCH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:07:52 2000; MasPar time 28.47 Seconds
Tabular output not generated. 33.572 Million cell updates/sec

Title: >US-09-362-731-5

Description: (1-32) from US09362731.pep

Perfect Score: 241

Sequence: 1 QYKANSKFIGITELGCGHGPCNIHRKPF 32

Scoring table: PAM 150

Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:Swissprot

Statistics: Mean 31.137; Variance 37.361; scale 0.833

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	122	50.6	146	1	DER2_DERPT	6.89e-15
2	122	50.6	146	1	DEF2_DERFA	6.89e-15
3	104	43.2	1314	1	TETX_CLOTE	4.44e-10
4	71	29.5	713	1	DDX4_RAT	3.47e-02
5	69	28.6	141	1	ALG2_TYRPU	9.19e-02
6	69	28.6	358	1	MET2_HAEIN	9.19e-02
7	67	27.8	141	1	LEP1_LEPDS	2.38e-01
8	66	27.4	149	1	EPI1_CANFA	3.81e-01
9	65	27.0	151	1	EPI1_HUMAN	6.06e-01
10	64	26.6	149	1	EPI1_BOVIN	9.59e-01
11	64	26.6	271	1	MURI_MYCTU	9.59e-01
12	64	26.6	333	1	PROBABLE GLUTAMATE RAC	9.59e-01
13	64	26.6	378	1	YD40_YEAST	9.59e-01
14	63	26.1	213	1	KAD_MYCCA	1.51e+00
15	63	26.1	261	1	EIA_ADE07	1.51e+00
16	63	26.1	899	1	V120_HSVSA	1.51e+00
17	62	25.7	284	1	RIBF_SYNY3	2.36e+00
18	62	25.7	287	1	TRUB_AQUAE	2.36e+00
19	62	25.7	339	1	TCMO_STRGA	2.36e+00
20	62	25.7	398	1	RFE_MYCLE	2.36e+00
21	62	25.7	404	1	RFE_MYCTU	2.36e+00
22	62	25.7	608	1	RDPO_SCEOB	2.36e+00
23	62	25.7	2156	1	ORP1_HUMAN	2.36e+00

Query Match 50.6%; Score 122; DB 1; Length 146;
Best Local Similarity 93.8%; Pred. No. 6.89e-15;

ALIGNMENTS

RESULT	ID	DER2_DERPT	STANDARD;	PRT;	146 AA.
AC	R49278;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).				
GN	DERP2.				
OS	Dermatophagoides pteronyssinus (House-dust mite).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;				
OC	Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;				
OC	Dermatophagoides.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 90256301.				
RA	CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,				
RA	THOMAS W.R.;				
RT	"Isolation of cDNA coding for the major mite allergen Der p II by IgE				
RT	plaque immunoassay."				
RL	Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).				
RN	[2]				
RP	PARTIAL SEQUENCE OF 18-57.				
RX	MEDLINE; 89278484.				
RA	HEYMAN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,				
RA	PLATTIS-MILLS T.A.;				
RT	"Antigenic and structural analysis of group II allergens (Der f II				
RT	and Der p II) from house dust mites (Dermatophagoides spp).";				
RL	J. Allergy Clin. Immunol. 83:1055-1067(1989).				
RN	[3]				
RP	STRUCTURE BY NMR.				
RX	MEDLINE; 98409423.				
RA	MUELLER G.A., BENJAMIN D.C., RULE G.S.;				
RT	"Tertiary structure of the major house dust mite allergen Der p 2:				
RT	sequential and structural homologies.";				
RL	Biochemistry 37:12707-12714(1998).				
CC	-1- SIMILARITY: TO MITE ALLERGEN LEP D I.				
DR	PDB; 1ASV; 14-OCT-98.				
KW	Allergen; Signal; 3D-structure.				
FT	SIGNAL	1	17		
FT	CHAIN	18	146		MITE ALLERGEN DER P 2.
FT	DISULFID	25	136		
FT	DISULFID	38	44		
FT	DISULFID	90	95		
SQ	SEQUENCE	146 AA;	15999 MW;	09A45F2E CRC32;	

Query Match 50.6%; Score 122; DB 1; Length 146;
Best Local Similarity 93.8%; Pred. No. 6.89e-15;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 37 GCGSEPCIIHRGKPF 52
 11111111 11111111

Qy 17 GCGSEPCIIHRGKPF 32

RESULT 2

ID DEF2_DERFA STANDARD; PRT; 146 AA.
 AC Q00855; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERRF2.
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcophagiformes; Astigmata; Pyroglyphidae; Pyroglyphidae;
 OC Dermatophagoides.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 91291341.
 RA YUKI T., OKUMURA Y., ANDO T., YAKAWA H., SUKO M., HAIDA M.,
 RA OKUDAIRA H.;
 RT "Cloning and expression of cDNA coding for the major house dust mite
 RT allergen Der f II in *Escherichia coli*.";
 RL Agric. Biol. Chem. 55:1233-1238(1991).
 RN [2]
 RP SEQUENCE OF 4-146 FROM N.A.
 RX MEDLINE: 94256850.
 RA OKUHARA H.;
 RT "Molecular biology of mite antigens.";
 RL Alerugi 43:435-440(1994).
 RN [3]
 RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE: 93283958.
 RA NISHIYAMA C., YUUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
 RT "Determination of three disulfide bonds in a major house dust mite
 RT allergen, Der f II.";
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).
 RN [4]
 RP PARTIAL SEQUENCE OF 18-52.
 RX MEDLINE: 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATTS-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE: 98079068.
 RA ICHIKAWA S., HATANAKA H., YUUKI T., IWAMOTO N., KOJIMA S.,
 RA NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
 RT "Solution structure of Der f 2, the major mite allergen for atopic
 RT diseases.";
 RL J. Biol. Chem. 273:356-360(1998).
 CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10447; BAA01239.1;
 CC EMBL: D10448; BAA01240.1;
 CC EMBL: D10449; BAA01241.1;
 CC EMBL: S70378; AAB30829.1;

DR PIR: PS0417; PS0417.
 DR PDB: 1AHK; 08-APR-98.
 DR PDB: 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 FT VARIANT 93 93 M -> V (IN CLONE 1).
 FT VARIANT 105 105 I -> A (IN CLONE 11).
 FT VARIANT 128 128 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
 SQ SEQUENCE 146 AA; 15802 MW; 72623F23 CRC32;

Query Match 50.6%; Score 122; DB 1; Length 146;
 Best Local Similarity 77.8%; Pred. No. 6.89e-15; Indels 0; Gaps 0;
 Matches 14; Conservative 2; Mismatches 2;

Db 35 VDCGSGSDPCIIHRGKPF 52
 : 11111111 11111111
 Qy 15 LGGCHGSEPCIIHRGKPF 32

RESULT 3

ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: primary structure, expression in *E. coli*, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911;
 RX MEDLINE: 87040747.
 RA FAIRWEATHER N.F., LYNESS V.A.;
 RT "The complete nucleotide sequence of tetanus toxin.";
 RL Nucleic Acids Res. 14:7809-7812(1986).
 RN [3]
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE: 86085672.
 RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in *Escherichia coli*.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE: 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE: 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]

RP IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc."
 RL EMBO J. 11:3577-3583(1992).
 [7]
 RN IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin."
 RL Nature 359:832-835(1992).
 [8]
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.
 RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin."
 RL Nat. Struct. Biol. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
 CC GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04436; CAA28033.1; -
 DR EMBL; M12739; AAA23282.1; -
 DR EMBL; X06214; CAA29564.1; -
 DR PIR; A25689; BYCLIN.
 DR PDB; 1AF9; 29-APR-98.
 DR PDB; 1ABD; 14-OCT-98.
 DR PROSITE; PS00142; ZINC-PROTEASE; 1.
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
 KW 3D-structure.
 FT INIT-MET 0 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT ACT 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 233 233 BY SIMILARITY.
 FT ACT_SITE 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 236 236 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT TRANSFID 438 466 INTERCHAIN.
 FT DISULFID 1076 1092
 FT SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
 Query Match 43.28; Score 104; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 4.44e-10;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYIKANSKFIGITEL 843
 QY 1 QYIKANSKFIGITEL 15
 RESULT 4
 ID DDX4_RAT STANDARD; PRT; 713 AA.
 AC Q64060;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DEAD BOX PROTEIN 4 (VASA HOMOLOG) (RVLG).
 GN DDX4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WISTAR; TISSUE-TESTIS;
 RX MEDLINE; 95160706.
 RA KOMIYA T., TANIGAWA Y.;
 RT "Cloning of a gene of the DEAD box protein family which is
 RT specifically expressed in germ cells in rats."
 RL Biochem. Biophys. Res. Commun. 207:405-410(1995).
 CC -!- TISSUE SPECIFICITY: TESTIS.
 CC -!- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES. HIGHEST TO
 CC DROSOPHILA VASA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S75275; AAB33364.1; -
 DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
 DR PFAM; PF00270; DEAD; 1.
 DR PFAM; PF00271; Helicase_C; 1.
 KW ATP-binding; Helicase; RNA-binding.
 FT NP_BIND 317 324 ATP (POTENTIAL).
 FT SITE 431 434 DEAD BOX.
 SQ SEQUENCE 713 AA; 77955 MW; 34CEB6D3 CRC32;
 Query Match 29.5%; Score 71; DB 1; Length 713;
 Best Local Similarity 24.2%; Pred. No. 3.47e-02;
 Matches 8; Conservative 13; Mismatches 11; Indels 1; Gaps 1;
 Db 479 EFLKSNLYFVAVGVGGACRDVQSQILOVGPVF 511
 QY 1 QYIKANSKFIGITELGG-CHGSEPCNIHRKPF 32
 RESULT 5
 ID ALG2_TYRPU STANDARD; PRT; 141 AA.
 AC O02380;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GROUP 2 ALLERGEN PRECURSOR.
 OS Tyrophagus putrescentiae (Dust mite)
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcotiformes; Astigmata; Acaroidea; Acaridae;
 OC Tyrophagus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC ERIKSSON T., JOHANSSON E., WHITLEY P., SCHMIDT M., ELSAYED S.,
 RA VAN HAGE-HAMSTEN M.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE E1 FAMILY.

FT REPEAT 64 65
 FT REPEAT 68 69
 FT REPEAT 72 73
 FT DISULFID 24 133
 FT DISULFID 37 42
 FT DISULFID 88 93
 FT VARIANT 35 35
 FT VARIANT 48 48
 FT VARIANT 53 53
 FT VARIANT 63 63
 FT VARIANT 90 91
 FT VARIANT 95 95
 FT VARIANT 104 104
 FT VARIANT 106 107
 FT VARIANT 116 116
 FT VARIANT 125 125
 FT VARIANT 136 136
 FT CONFLICT 26 26
 FT CONFLICT 30 30
 SQ SEQUENCE 141 AA; 14773 MW; 286421C CRC32;

Query Match 27.8%; Score 67; DB 1; Length 141;
 Best Local Similarity 50.0%; Pred. No. 2.38e-01;
 Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 33 DITGCGG-DTCVIHRG 47
 : : | | : | | | |
 QY 14 ELGGCHGSEPCNIHRG 29

RESULT 8
 ID EPI_CANFA STANDARD; PRT: 149 AA.
 AC Q28895;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE EPIDIDYMAL SECRETORY PROTEIN EI PRECURSOR (CEI).
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EPIDIDYMIS;
 RX MEDLINE; 95263175.
 RA ELLERBROCK K., PERA I., HARTUNG S., IVELL R.;
 RT "Gene expression in the dog epididymis: a model for human epididymal function";
 RL Int. J. Androl. 17:314-323(1994).
 CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: EPIDIDYMIS.
 CC -!- SIMILARITY: BELONGS TO THE EI FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S77411; AAB34263.1; -;
 DR SIGNAL
 FT CHAIN 1 21 POTENTIAL.
 FT DISULFID 22 149 EPIDIDYMAL SECRETORY PROTEIN EI.
 FT DISULFID 27 40 BY SIMILARITY.
 FT DISULFID 42 47 BY SIMILARITY.
 FT DISULFID 93 99 BY SIMILARITY.
 FT CARBOHYD 58 58 POTENTIAL.
 SQ SEQUENCE 149 AA; 16056 MW; 2908708D CRC32;

Query Match 27.4%; Score 66; DB 1; Length 149;
 Best Local Similarity 33.3%; Pred. No. 3.81e-01;
 Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 87 IPEADGCKSGINCPIQDKTY 107
 : : | | : | | : | :
 QY 12 ITELGGCHGSEPCNIHRGKDF 32

RESULT 9
 ID EPI_HUMAN STANDARD; PRT: 151 AA.
 AC Q15668; Q29413;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE EPIDIDYMAL SECRETORY PROTEIN EI PRECURSOR (EPI-1) (HE1) (EPIDIDYMAL SECRETORY PROTEIN 14.6) (ESPI4.6).
 OS Homo sapiens (Human), Pan troglodytes (Chimpanzee), and Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=HUMAN; TISSUE=EPIDIDYMIS;
 RX MEDLINE; 93119659.
 RA KRULL N., IVELL R., OSTERHOFF C., KIRCHHOFF C.;
 RT "Region-specific variation of gene expression in the human epididymis as revealed by in situ hybridization with tissue-specific cDNAs";
 RL Mol. Reprod. Dev. 34:16-24(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=P. FROGLDYTES; TISSUE=EPIDIDYMIS;
 RA FROHLICH O., YOUNG L.G., MURPHY T.C.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M. FASCICULARIS; TISSUE=EPIDIDYMIS;
 RX MEDLINE; 95180740.
 RA PERRY A.C.F., JONES R., HALL L.;
 RT "The monkey ESPI4.6 mRNA, a novel transcript expressed at high levels in the epididymis";
 RL Gene 153:291-292(1995).
 CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: EPIDIDYMIS.
 CC -!- SIMILARITY: BELONGS TO THE EI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X67698; CAA47928.1; -;
 DR EMBL; A18921; CAA01431.1; -;
 DR EMBL; U25748; AAA67077.1; -;
 DR EMBL; X78134; CAA55013.1; -;
 DR SIGNAL
 FT CHAIN 1 19 POTENTIAL.
 FT DISULFID 20 151 EPIDIDYMAL SECRETORY PROTEIN EI.
 FT DISULFID 27 140 BY SIMILARITY.
 FT DISULFID 42 47 BY SIMILARITY.
 FT DISULFID 93 99 BY SIMILARITY.
 FT CARBOHYD 58 58 POTENTIAL.
 SQ SEQUENCE 151 AA; 16570 MW; 2EC86F5 CRC32;

Query Match 27.0%; Score 65; DB 1; Length 151;
 Best Local Similarity 33.3%; Pred. No. 6.06e-01;
 Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 87 IPEPDGCKSGINCPIQDKTY 107
 : : | | : | | : | :
 QY 12 ITELGGCHGSEPCNIHRGKDF 32

RESULT 10
ID EPL_BOVIN STANDARD; PRT; 149 AA.
AC P79345;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SECRETORY PROTEIN EI PRECURSOR (EPV20).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RC TISSUE-MAMMARY GLAND;
RX MEDLINE; 97182631.
RA LARSEN L.B., RAVN P., BOISEN A., BERGLUND L., PETERSEN T.E.;
RT "Primary structure of EPV20, a secretory glycoprotein containing a previously uncharacterized type of domain."
RL Eur. J. Biochem. 243:437-441(1997).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY, SPLEEN, LIVER AND MAMMARY GLAND, BUT NOT IN TESTIS.
CC -1- SIMILARITY: BELONGS TO THE EI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X85799; CAA59794.1;
KW Signal.
FT SIGNAL 1 19
FT CHAIN 20 149 SECRETORY PROTEIN EI.
FT DISULFID 27 140
FT DISULFID 42 47
FT DISULFID 93 99
FT CARBOHYD 58 58 POTENTIAL.
SQ SEQUENCE 149 AA; 16640 MW; AFF60F1 CRC32;

Query Match 26.6%; Score 64; DB 1; Length 149;
Best Local Similarity 31.6%; Pred. No. 9.59e-01;
Matches 6; Conservative 9; Mismatches 3; Indels 1; Gaps 1;

Db 38 NVSPCP-TQCKLHGQSY 55
::: | :|:|:|:| :|:
QY 14 ELGGCHGSEPCNTHRGKPF 32

RESULT 11
ID MURI_MYCTU STANDARD; PRT; 271 AA.
AC Q10626;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 39, Last annotation update)
DE PROBABLE GLUTAMATE RACEMASE (EC 5.1.1.3).
GN MURI OR RV1338 OR MTCY130.23 OR MTCY02B10.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAWLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,

RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: PROVIDES THE (R)-GLUTAMIC ACID REQUIRED FOR CELL WALL BIOSYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE - D-GLUTAMATE.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ASPARTATE/GLUTAMATE RACEMASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; Z73902; CAA98102.1;
DR PROSITE; PS00923; ASP_GLU_RACEMASE_1; 1.
DR PROSITE; PS00924; ASP_GLU_RACEMASE_2; 1.
DR PFAM; PF01177; ASP_Glu_race; 1.
KW Peptidoglycan synthesis; Cell wall; Isomerase.
SQ SEQUENCE 271 AA; 28643 MW; 2219C126 CRC32;

Query Match 26.6%; Score 64; DB 1; Length 271;
Best Local Similarity 34.6%; Pred. No. 9.59e-01;
Matches 9; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 247 FTKLAARFLG-PVLGGVQVPHPSRIH 271
: | :|:|:|:| :|:
QY 2 YIRANSFTEIGTELGCGHSEPCNIH 27

RESULT 12
ID DPOB_XENLA STANDARD; PRT; 333 AA.
AC Q57383;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE BETA (EC 2.7.7.7).
GN POLB.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98151235.
RA REICHENBERGER S., PFEIFFER P.;
RT "Cloning, purification and characterization of DNA polymerase beta from Xenopus laevis -- studies on its potential role in DNA-end joining."
RL Eur. J. Biochem. 251:81-90(1998).
CC -1- FUNCTION: REPAIR POLYMERASE. CONDUCTS "GAP-FILLING" DNA SYNTHESIS IN A STEPWISE DISTRIBUTIVE FASHION RATHER THAN IN A PROCESSIVE FASHION AS FOR OTHER DNA POLYMERASES.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES: ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; Y15732; CAA75741.1; -
DR PROSITE; PS00522; DNA_POLYMERASE_X; 1.
DR PFAM; PF00966; DNA_polymeraseX; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 182 182 INVOLVED IN PRIMER BINDING
FT ACT_SITE 189 189 (BY SIMILARITY).
FT ACT_SITE 189 189 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 (BY SIMILARITY).
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 (BY SIMILARITY).
SQ SEQUENCE 333 AA; 38162 MW; 6948CECC CRC32;

Query Match 26.6%; Score 64; DB 1; Length 333;
Best Local Similarity 26.9%; Pred. No. 9.59e-01;
Matches 7; Conservative 10; Mismatches 8; Indels 1; Gaps 1;

Db 228 VKGDKFMGVCLP-CESDQDPYVRR 252
QY 3 IKANSKFIGITELGGCHGSEPCNIHR 28

RESULT 13
ID YD40.YEAST STANDARD; PRT; 378 AA.
AC Q04179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 42 3 KD PROTEIN IN YTA2-DIT1 INTERGENIC REGION.
GN YD400W OR D9509.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
RA BERNO A., CARPENTER J., CHEN E., CHERRY J.M., CHUNG E., DUNCAN M.,
RA HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAWATH A., OEFNER P., OH C., PETEL F.X.,
RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGREN T., SHROFF N.,
RA WINANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE EMBL/GENBANK/DBJ DATABASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06414; CAA29724.1; -
DR PIR; S02851; KIYMC.
DR HSSP; P27142; 12IO.
DR PROSITE; PS00113; ADENYLATE KINASE; 1.
DR PFAM; PF00406; adenylatekinase; 1.
KW Transferase; Kinase; ATP-binding.
FT NP_BIND 7 15 ATP (BY SIMILARITY).
SQ SEQUENCE 213 AA; 24616 MW; 8D5C4B03 CRC32;

Query Match 26.1%; Score 63; DB 1; Length 213;
Best Local Similarity 66.7%; Pred. No. 1.51e-00;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 182 DYFKTNSKFIEI 193
QY 1 QVIKANSKFIGI 12

RESULT 15
ID E1A.ADR07 STANDARD; PRT; 261 AA.
AC P03256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EARLY E1A 28 KD PROTEIN [CONTAINS: EARLY E1A 24 KD PROTEIN].
OS Human adenovirus type 7.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GONEN;
RX MEDLINE; 81237792.
RA DIJKEMA R., DEKKER B.M.M., VAN ORMONDT H., DE WAARD A., MAAT J.,
RA BOYER H.W.;
RT "Gene organization of the transforming region of weakly oncogenic
RT adenovirus type 7: the E1a region.";
RL Gene 12:287-299(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-GRIDER;
RA YOSHIDA K., FUJINAGA K.;
RT "The nucleotide sequence of the transforming HindIII-I.J fragment of
RT adenovirus type 7 DNA.";
RL Tumor Res. 19:39-47(1984).
```


(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Result No.	Query %		DB	ID	Description	Pred. No.
	Score	Match				
1	122	50.6	129	2	A61501	allergen Der f II - h 1.40e-12
2	122	50.6	129	2	JU00394	allergen Der f II (pF 1.40e-12
3	122	50.6	138	2	A61241	allergen Der f II pre 1.40e-12
4	122	50.6	138	2	B61241	allergen Der f II pre 1.40e-12
5	122	50.6	146	2	A60381	allergen Der p II pre 1.40e-12
6	104	43.2	1315	1	BTCLTN	tenutoxylisin (EC 3.4. 2.26e-08
7	75	31.1	533	2	A47143	retinal pigment micro 3.39e-02
8	72	29.9	173	2	T00970	hypothetical protein 1.28e-01
9	71	29.5	713	2	JC2534	RVLG protein - rat 1.99e-01
10	71	29.5	867	2	T14777	hypothetical protein 1.99e-01
11	69	28.6	102	2	T16978	ribosomal protein Sl7 4.71e-01
12	69	28.6	141	2	B66499	allergen Lep d 1.02 p 4.71e-01
13	69	28.6	358	2	D64113	homoserine o-acetyltr 4.71e-01
14	69	28.6	502	2	T05135	hypothetical protein 4.71e-01
15	68	28.2	1036	2	B69368	hypothetical protein 7.20e-01
16	67	27.8	141	2	S66500	allergen Lep d 1.01 p 1.10e+00
17	66	27.4	149	2	I69229	CE1 - dog 1.66e+00
18	66	27.4	379	2	D70846	probable meta protein 1.66e+00
19	65	27.0	151	2	I53929	epididymal secretory 2.51e+00
20	65	27.0	151	2	I38465	epididymal secretory 2.51e+00
21	64	26.6	38	2	G64583	hypothetical protein 3.78e+00
22	64	26.6	159	2	A72660	hypothetical protein 3.78e+00
23	64	26.6	271	2	F70771	probable glutamaterac 3.78e+00


```

CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 50.6%; Score 122; DB 2; Length 129;
Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 18 VDCGSGDPCIIHRGKPF 35
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||:| |||||

RESULT 3
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS A61241
REFERENCE A61241; PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession A61241
#molecule_type mRNA
#residues 1-138 #label YUO
#note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
10-138 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status experimental #label
MAT
SUMMARY #length 138 #checksum 2894

Query Match 50.6%; Score 122; DB 2; Length 138;
Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 27 VDCGSGDPCIIHRGKPF 44
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||:| |||||

RESULT 4
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS B61241; JU0395
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession B61241
#molecule_type mRNA
#residues 1-138 #label YUO
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
10-138 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 50.6%; Score 122; DB 2; Length 138;

Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 27 VDCGSGDPCIIHRGKPF 44
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||:| |||||

RESULT 5
ENTRY A60381 #type complete
TITLE allergen Der p II precursor - house-dust mite
(Dermatophagoides pteronyssinus)
ORGANISM #formal_name Dermatophagoides pteronyssinus
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A60381
REFERENCE A60381
#authors Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;
Stewart, G.A.; Thomas, W.R.
#journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
#title Isolation of cDNA coding for the major mite allergen Der p II
by IgE plaque immunoassay.
#cross-references MUID:90256301
#accession A60381
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-146 #label CHU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-17
18-146 #domain signal sequence #status predicted #label SIG\
#product allergen Der p II #status predicted #label MAT
SUMMARY #length 146 #molecular-weight 15999 #checksum 25

Query Match 50.6%; Score 122; DB 2; Length 146;
Best Local Similarity 93.8%; Pred. No. 1.40e-12;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 37 GCHGSEPCNIHRGKPF 52
QY 17 GCHGSEPCNIHRGKPF 32
: |||||:| |||||

RESULT 6
ENTRY BTCLTN #type complete
TITLE tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani
tetanus neurotoxin
ALTERNATE_NAMES #formal_name Clostridium tetani
ORGANISM 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
18-Jun-1999
ACCESSIONS A25689; A25757; A25194; B25194; A60759; S69348; S09364
REFERENCE A25689
#authors Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels,
J.; Weiler, U.; Hudel, M.; Habermann, E.; Niemann, H.
#journal EMBO J. (1986) 5:2495-2502
#title Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
#molecule_type DNA
#residues 1-1315 #label EIS
#cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
#molecule_type DNA
#residues 1-1315 #label FAI
#cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
#experimental_source strain CN3911
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.;

```

```

#journal      Thomson, R.O.
#title        J. Bacteriol. (1986) 165:21-27
#cross-references MUID:86085672
#accession    A25194
#molecule_type DNA
#residues     743-1315 #label FA2
#cross-references GB:M12739; NID:G144920; PIDN:AAA23282.1; PID:G144921
#accession    B25194
#molecule_type protein
#residues     865-894 #label FA3
REFERENCE
#authors      Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
#journal      Infect. Immun. (1989) 57:3588-3593
#title        Isolation, purification, and characterization of fragment B,
              the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession    A60759
#molecule_type protein
#residues     461-475 #label MAT
REFERENCE
#authors      Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.;
              Widmann, C.; Corradin, G.
#journal      J. Immunol. (1989) 142:394-402
#title        Delineation of several DR-restricted tetanus toxin T cell
              epitopes.
#cross-references MUID:89093918
#contents     annotation; epitope region
REFERENCE
#authors      Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
              Laureto, P.P.; Dasgupta, B.R.; Montecucco, C.
#journal      Nature (1992) 359:832-835
#title        Tetanus and botulinum-B neurotoxins block neurotransmitter
              release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents     annotation
REFERENCE
#authors      de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.;
              Montecucco, C.
#journal      Eur. J. Biochem. (1995) 229:61-69
#title        Structural studies on the zinc-endoropeptidase light chain of
              tetanus neurotoxin.
#cross-references MUID:95262688
#accession    S69348
#molecule_type protein
#residues     2-31 #label DEF
COMMENT
#description  The source of this protein was an extrachromosomal plasmid.
              The precursor is cleaved by endogenous proteinase activity to form
              light (fragment A) and heavy (fragment B.C) chains that are
              covalently linked by an interchain disulfide bond (the individual
              chains are not toxic when separated). The amino end of the heavy
              chain (fragment B) can be separated from the carboxyl end
              (fragment C) by papain.
COMMENT
#description  Fragment B forms ion channels in a lipid bilayer. Fragment C binds
              to gangliosides and may target the toxin to the motor end plate.
              Fragment A is a zinc-dependent endopeptidase.
COMMENT
#description  This potent neurotoxin binds to peripheral neuronal synapses, is
              internalized, and moves by retrograde transport up the axon into
              the spinal cord where it can move between postsynaptic and
              presynaptic neurons. It inhibits neurotransmitter release by
              proteolytic cleavage of synaptobrevin (vesicle-associated
              membrane protein 2).
FUNCTION
#description  blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide
              bond in synaptobrevin 2
CLASSIFICATION
#superfamily  tetanus toxin
KEYWORDS
#keywords     hydrolase; metalloproteinase; neurotoxin; transmembrane
              protein; zinc
FEATURE
#feature      #product tentoxylisin light chain (fragment A) #status
              predicted #label TTL\
2-457        #product tentoxylisin heavy chain (fragment B.C) #status
              predicted #label TTL\
461-1315

```

```

experimental #label TTH\
#domain channel forming (fragment B) #status predicted
#label TXB\
#domain ganglioside binding (fragment C) #status
865-1315    predicted #label TXC\
#binding_site zinc (His) #status predicted\
233-237    #active_site Glu #status predicted
234        #length 1315 #molecular-weight 150681 #checksum 4853
SUMMARY
Query Match      43.2%; Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 2.26e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 830 QYIKANSKFIGITEL 844
QY 1 QYIKANSKFIGITEL 15
|||||
RESULT 7
ENTRY A47143 #type complete
TITLE retinal pigment microsomal protein RPE65, epithelium-specific
ALTERNATE_NAMES - bovine
ORGANISM membrane receptor p63; retinol-binding protein receptor
#formal_name Bos primigenius taurus #common_name cattle
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
ACCESSIONS A47143; A48017; S28503
REFERENCE A47143
#authors Hamel, C.P.; Tsilou, E.; Pfeiffer, B.A.; Hooks, J.J.; Detrick,
              B.; Redmond, T.M.
#journal J. Biol. Chem. (1993) 268:15751-15757
#title Molecular cloning and expression of RPE65, a novel retinal
              pigment epithelium-specific microsomal protein that is
              post-transcriptionally regulated in vitro.
#cross-references MUID:93340181
#accession A47143 preliminary
#status ##molecule_type mRNA
#residues 1-533 #label HAM
#cross-references GB:L11356; NID:G163656; PID:G163657
REFERENCE A48017
#authors Baviak, C.O.; Levy, F.; Hellman, U.; Wernstedt, C.; Eriksson,
              U.
#journal J. Biol. Chem. (1993) 268:20540-20546
#title The retinal pigment epithelial membrane receptor for plasma
              retinol-binding protein. Isolation and cDNA cloning of the
              63-kDa protein.
#cross-references MUID:93388633
#accession A48017 preliminary
#status ##molecule_type mRNA
#residues 1-340, 'L', 342-533 #label BA2
#cross-references EMBL:X66277; NID:G563; PID:G564
KEYWORDS
#keywords     membrane protein
SUMMARY
#length 533 #molecular-weight 60944 #checksum 3885
Query Match      31.1%; Score 75; DB 2; Length 533;
Best Local Similarity 44.0%; Pred. No. 3.39e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
QY 2 YIKANS-KFIGITELGGCHGSEPCN 25
|||||
RESULT 8
ENTRY T00970 #type complete
TITLE hypothetical protein T9J22.4 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
              cross
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
20-Sep-1999
ACCESSIONS T00970

```

```

REFERENCE      Z14153
#authors       Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
               Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
               Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
               J.C.
#description   submitted to the EMBL Data Library, April 1998
#description   Arabidopsis thaliana chromosome II BAC T9J22 genomic
               sequence.
#accession     T00970
#status        translated from GB/EMBL/DBJ
#molecule_type DNA
#residues       1-173 #label ROU
#cross-references EMBL:AC002505; NID:g2739359; PID:g2739363
#experimental_source cultivar Columbia

GENETICS
#map_position  2
#introns       34/1; 74/1; 132/3
#note          T9J22.4
CLASSIFICATION #superfamily Arabidopsis thaliana hypothetical protein
               T9J22.4
SUMMARY        #length 173 #molecular-weight 19015 #checksum 2215
               29.9%; Score 72; DB 2; Length 173;
Query Match    34/1; 74/1; 132/3
Best Local Similarity 35.5%; Pred. No. 1.28e-01;
Matches 11; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

Db 88 YENVNSTIINY-DLSDVMGEDPCSIPEGKF 117
      | : | | | : | : | : | | | | |
QY 2 YIKANSKFGITELGGCHGSEPCNIHRGKPF 32
      | : | | | : | : | : | | | | |

RESULT 9
ENTRY   JC2534 #type complete
TITLE   RVLG protein - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE    20-May-1995 #sequence_revision 14-Jul-1995 #text_change
        20-Sep-1999
ACCESSIONS JC2534
REFERENCE   JC2534
#authors    Komiya, T.; Tanigawa, Y.
#journal     Blochem. Biophys. Res. Commun. (1995) 207:405-410
#title       Cloning of a gene of the DEAD Box protein family which is
               specifically expressed in germ cells in rats.
#cross-references MUID:95160706
#accession     JC2534
#molecule_type mRNA
#residues       1-713 #label KOM
#cross-references GB:S75275; NID:g806463; PIDN:AA33364.1; PID:g806464
COMMENT      This protein contains a conserved DEAD box.
CLASSIFICATION #superfamily unassigned DEAD/H box helicases; DEAD/H box
               helicase homology
KEYWORDS      ATP; P-loop
FEATURE
317-618      #domain DEAD/H box helicase homology #label DEAD\
317-324      #region nucleotide-binding motif A (P-loop)\
427-432      #region nucleotide-binding motif B\
431-434      #region DEAD motif
SUMMARY        #length 713 #molecular-weight 77954 #checksum 5585
               29.5%; Score 71; DB 2; Length 713;
Query Match    29.5%; Score 71; DB 2; Length 713;
Best Local Similarity 24.2%; Pred. No. 1.99e-01;
Matches 8; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

Db 479 EFLKSNLYFVAGVGACGRDVOQSIIQVGVF 511
      : : | | | : : | | : : | |
QY 1 QYIKANSKFGITELGG-CHGSEPCNIHRGKPF 32
      : : | | | : : | | : : | |

RESULT 10
ENTRY   T14777 #type fragment
TITLE   hypothetical protein DKF2p434N061.1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE    20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change

```

```

ACCESSIONS     20-Sep-1999
REFERENCE       T14777
#authors       Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.;
               Wiemann, S.
#submission    submitted to the Protein Sequence Database, August 1999
#accession     T14777
#status        preliminary
#molecule_type mRNA
#residues       1-867 #label POU
#cross-references EMBL:AL110249
#experimental_source adult testis; clone DKF2p434N061
GENETICS
#note          DKF2p434N061.1
SUMMARY        #length 867 #checksum 3074
               29.5%; Score 71; DB 2; Length 867;
Query Match    29.5%; Score 71; DB 2; Length 867;
Best Local Similarity 44.0%; Pred. No. 1.99e-01;
Matches 11; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

Db 432 IESDKVIISIVTGLPGCHASELCAF 456
      | : | | | : | : | : | | | | |
QY 3 IKANSKFIGI-TELGCGHSEPCNI 26
      | : | | | : | : | : | | | | |

RESULT 11
ENTRY   T16978 #type complete
TITLE   ribosomal protein S17 - curled-leaved tobacco
ORGANISM #formal_name Nicotiana glumabaghiifolia #common_name
               curled-leaved tobacco
DATE    20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
        20-Sep-1999
ACCESSIONS     T16978
REFERENCE       T16978
#authors       Borisjuk, N.V.
#submission    submitted to the EMBL Data Library, October 1996
#accession     T16978
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues       1-102 #label BOR
#cross-references EMBL:F08858; NID:e1008001; PID:e275056
               protein biosynthesis; ribosome
SUMMARY        #length 102 #molecular-weight 11602 #checksum 6192
               28.6%; Score 69; DB 2; Length 102;
Query Match    28.6%; Score 69; DB 2; Length 102;
Best Local Similarity 21.9%; Pred. No. 4.71e-01;
Matches 7; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Db 32 RYVKRTSKFMAHQNDQNCNIGDRVKLDHSRPL 63
      | : | | | : | : | : | | | | |
QY 1 QYIKANSKFGITELGGCHGSEPCNIHRGKPF 32
      | : | | | : | : | : | | | | |

RESULT 12
ENTRY   S66499 #type complete
TITLE   allergen Lep d 1.02 precursor - Lepidoglyphus destructor
ORGANISM #formal_name Lepidoglyphus destructor
DATE    19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
        29-Sep-1999
ACCESSIONS     S66499
REFERENCE       S66499
#authors       Schmidt, M.; Olsson, S.; van der Ploeg, I.; van Hage-Hamsten,
               M.
#journal       FEBS Lett. (1995) 370:11-14
#title         CDNA analysis of the mite allergen Lep d 1 identifies two
               different isoallergens and variants.
#cross-references MUID:95377437
#accession     S66499
#molecule_type mRNA
#residues       1-141 #label SCH
#cross-references EMBL:X83875; NID:g999457; PIDN:CAA58755.1;
               PID:g999458
CLASSIFICATION #superfamily allergen Der p II

```

```

FEATURE
1-16      #domain signal sequence #status predicted #label SIG
17-141    #product allergen Lep d 1.02 #status predicted #label
MAP
SUMMARY   #length 141 #molecular-weight 14812 #checksum 5305
Query Match      28.6%; Score 69; DB 2; Length 141;
Best Local Similarity 50.0%; Pred. No. 4.71e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 33 DISGCGS-DTCVIHRG 47
      ::::| | | | |
QY 14 ELGGCHGSEPCNIHRG 29

RESULT 13
ENTRY   #type complete
TITLE   homoserine O-acetyltransferase homolog - Haemophilus
ORGANISM influenzae (strain Rd KW20)
#formal_name Haemophilus influenzae
DATE    18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
24-Oct-1997
ACCESSIONS D64113
REFERENCE   A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kierlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal   Science (1995) 269:496-512
#title     Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
#cross-references MUID:195350830
#accession D64113
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  1-358 #label TIGR
#cross-references GB:U32806; GB:L42023; NID:gl574717; PID:gl574721;
TIGR:HL1263
SUMMARY   #length 358 #molecular-weight 39990 #checksum 6643
Query Match      28.6%; Score 69; DB 2; Length 358;
Best Local Similarity 44.4%; Pred. No. 4.71e-01;
Matches 12; Conservative 6; Mismatches 6; Indels 3; Gaps 3;

Db 82 FISSNLVGGCKGTGPSSINPQTGKPY 108
      ||: ||||| | | | | |
QY 9 FIGITELGGCHGS-EPCNIH-R-GRPF 32

RESULT 14
ENTRY   #type complete
TITLE   hypothetical protein F7H19.220 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE    23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
23-Jul-1999
ACCESSIONS T05135
REFERENCE   Z15399
#authors   Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.;
Stiekema, W.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.;
Schueller, C.
#submission submitted to the Protein Sequence Database, July 1998
#accession T05135
#molecule_type DNA
#residues  1-502 #label BEV
#cross-references EMBL:AL031018
#experimental_source cultivar Columbia; BAC clone F7H19
GENETICS
#map_position 4
#note         F7H19.220
SUMMARY   #length 502 #molecular-weight 55232 #checksum 5495
Query Match      28.6%; Score 69; DB 2; Length 502;
Best Local Similarity 42.9%; Pred. No. 4.71e-01;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 382 IIGLCELGNCPTTLCGLRG 402
      ::::| | | | |
QY 9 FIGITELGGCHGSEPCNIHRG 29

RESULT 15
ENTRY   #type complete
TITLE   hypothetical protein AF0946 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE    05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS B69368
REFERENCE   A69250
#authors   Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal   Nature (1997) 390:364-370
#title     The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession B69368
#status    preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues  1-1036 #label KLE
#cross-references GB:AE001039; GB:AE000782; NID:g3689362; PID:g3649670;
TIGR:AF0946
SUMMARY   #length 1036 #molecular-weight 112425 #checksum 562
Query Match      28.2%; Score 68; DB 2; Length 1036;
Best Local Similarity 30.0%; Pred. No. 7.20e-01;
Matches 9; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 543 INSEVFK-IDDLAFSGDPCDV-REKTY 570
      ::::| | | | |
QY 3 IKANSFIGITELGGCHGSEPCNIHRGKPF 32

Search completed: Sun Apr 2 18:07:34 2000
Job time : 49 secs.

```

[M][A][P][E][R][C][H] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:10:13 2000; MasPar time 2.25 Seconds
Tabular output not generated. 184,493 Million cell updates/sec

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QYIKANSKFITGELGCGSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 21.282; Variance 66.961; scale 0.318

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query Match	Length	Description	DB	ID
1	122	50.6	25	PCT-US95-0	Sequence 19, Applicati	6.81e-06
2	122	50.6	25	US-08-482-	Sequence 42, Applicati	6.81e-06
3	122	50.6	25	US-08-478-	Sequence 104, Applicat	6.81e-06
4	122	50.6	25	US-08-478-	Sequence 42, Applicati	6.81e-06
5	122	50.6	61	US-08-482-	Sequence 162, Applicat	6.81e-06
6	122	50.6	61	US-08-478-	Sequence 162, Applicat	6.81e-06
7	122	50.6	84	US-08-482-	Sequence 161, Applicat	6.81e-06
8	122	50.6	84	US-08-478-	Sequence 161, Applicat	6.81e-06
9	122	50.6	92	US-08-482-	Sequence 160, Applicat	6.81e-06
10	122	50.6	92	US-08-478-	Sequence 160, Applicat	6.81e-06
11	122	50.6	129	US-08-482-	Sequence 159, Applicat	6.81e-06
12	122	50.6	129	US-07-945-	Sequence 8, Applicatio	6.81e-06
13	122	50.6	129	US-08-462-	Sequence 8, Applicatio	6.81e-06
14	122	50.6	129	US-08-461-	Sequence 8, Applicatio	6.81e-06
15	122	50.6	129	US-08-462-	Sequence 12, Applicati	6.81e-06
16	122	50.6	129	US-08-482-	Sequence 8, Applicatio	6.81e-06
17	122	50.6	129	US-08-462-	Sequence 13, Applicati	6.81e-06
18	122	50.6	129	US-07-945-	Sequence 12, Applicati	6.81e-06
19	122	50.6	129	US-08-461-	Sequence 8, Applicatio	6.81e-06
20	122	50.6	129	PCT-US93-0	Sequence 12, Applicati	6.81e-06
21	122	50.6	129	PCT-US93-0	Sequence 13, Applicati	6.81e-06
22	122	50.6	129	PCT-US93-0	Sequence 8, Applicatio	6.81e-06
23	122	50.6	129	US-08-478-	Sequence 157, Applicat	6.81e-06

24	122	50.6	129	1	US-07-945-	Sequence 13, Applicati	6.81e-06
25	122	50.6	129	1	US-08-461-	Sequence 13, Applicati	6.81e-06
26	122	50.6	129	1	US-08-461-	Sequence 12, Applicati	6.81e-06
27	122	50.6	129	2	US-08-482-	Sequence 157, Applicat	6.81e-06
28	122	50.6	129	2	US-08-478-	Sequence 159, Applicat	6.81e-06
29	122	50.6	129	2	US-08-478-	Sequence 8, Applicatio	6.81e-06
30	122	50.6	129	1	US-08-461-	Sequence 13, Applicati	6.81e-06
31	122	50.6	129	1	US-08-461-	Sequence 12, Applicati	6.81e-06
32	122	50.6	130	2	US-08-482-	Sequence 158, Applicat	6.81e-06
33	122	50.6	130	2	US-08-478-	Sequence 158, Applicat	6.81e-06
34	122	50.6	142	1	US-08-288-	Sequence 4, Applicatio	6.81e-06
35	122	50.6	142	2	US-08-910-	Sequence 4, Applicatio	6.81e-06
36	122	50.6	142	2	US-08-910-	Sequence 4, Applicatio	6.81e-06
37	122	50.6	142	2	US-08-905-	Sequence 2, Applicatio	6.81e-06
38	122	50.6	142	2	US-08-905-	Sequence 4, Applicatio	6.81e-06
39	122	50.6	142	1	US-08-288-	Sequence 2, Applicatio	6.81e-06
40	122	50.6	146	1	US-08-462-	Sequence 4, Applicatio	6.81e-06
41	122	50.6	146	1	US-08-461-	Sequence 4, Applicatio	6.81e-06
42	122	50.6	146	2	US-08-482-	Sequence 4, Applicatio	6.81e-06
43	122	50.6	146	3	PCT-US93-0	Sequence 4, Applicatio	6.81e-06
44	122	50.6	146	1	US-07-945-	Sequence 4, Applicatio	6.81e-06
45	122	50.6	146	2	US-08-478-	Sequence 4, Applicatio	6.81e-06

ALIGNMENTS

RESULT 1 STANDARD; PRT; 25 AA.

ID PCT-US95-04481-19

XX xxxxxx

XX

XX

XX

XX

XX

XX

Sequence 19, Application PC/TUS9504481

CC

Sequence 19, Application PC/TUS9504481

CC

GENERAL INFORMATION:

CC APPLICANT:

CC TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Du

CC NUMBER OF SEQUENCES: 54

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US95/04481

CC FILING DATE:

CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/227,772

CC FILING DATE: April 14, 1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Vanstone, Darlene A.

CC REGISTRATION NUMBER: 35,279

CC REFERENCE/DOCKET NUMBER: 017.5 PCT

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000

CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 19:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 25 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 25 AA; 2782 MW; 3408 CN;

SQ

Query Match 50.6%; Score 122; DB 3; Length 25;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC	APPLICANT:	Greenstein, Julia
CC	APPLICANT:	Kuo, Mei-chang
CC	APPLICANT:	Rogers, Bruce
CC	APPLICANT:	Franzen, Henry
CC	APPLICANT:	Chen, Xian
CC	APPLICANT:	Evans, Sean
CC	APPLICANT:	Shaked, Ze'ev
CC	TITLE OF INVENTION:	T CELL EPITOPES OF THE MAJOR ALLERGENS
CC	TITLE OF INVENTION:	FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC	NUMBER OF SEQUENCES:	207
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESS:	IMMULOGIC PHARMACEUTICAL CORPORATION
CC	STREET:	610 LINCOLN STREET
CC	CITY:	WALTHAM
CC	STATE:	MA
CC	COUNTRY:	USA
CC	ZIP:	02154
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	ASCII TEXT
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/482,142
CC	FILING DATE:	07-JUN-1995
CC	CLASSIFICATION:	435
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/445,307
CC	FILING DATE:	07 June 1995
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	CRAIG, ANNE I.
CC	REGISTRATION NUMBER:	32,976
CC	REFERENCE/DOCKET NUMBER:	017.6US
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(617) 466-6000
CC	TELEFAX:	(617) 466-6040
CC	INFORMATION FOR SEQ ID NO:	162:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	61 amino acids
CC	TYPE:	amino acid
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	peptide
CC	FRAGMENT TYPE:	internal
SQ	SEQUENCE	61 AA; 6669 MW; 17328 CN;
	Query Match	50.68; Score 122; DB 2; Length 61;
	Best Local Similarity	93.8%; Pred.No. 6.81e-06;
	Matches	15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	20 GCHGSEPCIIHRGKPF 35	
Qy	17 GCHGSEPCNIHRGKPF 32	
RESULT	6	
ID	US-08-478-572-162	STANDARD; PRT; 61 AA.
XX	xxxxxx	
AC		
DT		
XX		
DE		
XX		
XX		
CC	Sequence 162, Application US/08478572	
CC	Sequence 162, Application US/08478572	
CC	Patent No. 5968356	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Garmen, Richard
CC	APPLICANT:	Greenstein, Julia
CC	APPLICANT:	Kuo, Mei-chang
CC	APPLICANT:	Rogers, Bruce
CC	APPLICANT:	Franzen, Henry
CC	APPLICANT:	Chen, Xian
CC	APPLICANT:	Evans, Sean

CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 61 AA; 6669 MW; 17328 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 61;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 7
ID US-08-482-142-161 STANDARD; PRT: 84 AA.
XX
AC xxxxxx
XX
DT
Sequence 161, Application US/08482142
Sequence 161, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:

CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 84;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 8
ID US-08-478-572-161 STANDARD; PRT: 84 AA.
XX
AC xxxxxx
XX
DT
Sequence 161, Application US/08478572
Sequence 161, Application US/08478572
Patent No. 5968526
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC
SQ SEQUENCE 84 AA; 9150 MW; 34277 CN;

Query Match 50.6%; Score 122; DB 2; Length 84;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
|||||
RESULT 9
ID US-08-482-142-160 STANDARD; PRT; 92 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 160, Application US/08482142
XX
CC Sequence 160, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142

CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 160:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 92 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: Internal
CC
SQ SEQUENCE 92 AA; 10222 MW; 43773 CN;

Query Match 50.6%; Score 122; DB 2; Length 92;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
|||||
RESULT 10
ID US-08-478-572-160 STANDARD; PRT; 92 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 160, Application US/08478572
XX
CC Sequence 160, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:

CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60S
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 160:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 92 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 92 AA; 10222 MW; 43773 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 92;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||||
RESULT 11
ID US-08-482-142-159 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 159, Application US/08482142
DE
XX
Sequence 159, Application US/08482142
CC
Patent No. 5820862
CC
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC
CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC
COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC
CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC
PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC
ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60S
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 129;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||||
RESULT 12
ID US-07-945-288-8 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 8, Application US/07945288
DE
XX
Sequence 8, Application US/07945288
CC
Patent No. 5433948
CC
GENERAL INFORMATION:
CC APPLICANT: Thomas, Wayne R.
CC APPLICANT: Chua, Kew-Yan
CC TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)
CC NUMBER OF SEQUENCES: 13
CC
CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC
COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC
CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/945,288
CC FILING DATE: 19920910
CC CLASSIFICATION: 514
CC
PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 580,655
CC FILING DATE: 11 SEPTEMBER 1990
CC
APPLICATION NUMBER: 458,642
CC FILING DATE: 13 FEBRUARY 1990
CC
ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC
INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 129 AA; 14022 MW; 84331 CN;
SQ
Query Match 50.6%; Score 122; DB 1; Length 129;
Best Local Similarity 77.8%; Pred. No. 6.81e-06;

Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 18 VDCGSGSDPCIHRGKPF 35
: |||||:|||||
QY 15 LGGCHGSEPCNIHRGKPF 32

RESULT 13
ID US-08-462-831-8 STANDARD; PRT; 129 AA.
XX
XX
XX
DT
XX
DE
XX
Sequence 8, Application US/08462831
Sequence 8, Application US/08462831
Patent No. 552142
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,831
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 129 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 129 AA; 14022 MW; 84331 CN;

Query Match 50.6%; Score 122; DB 1; Length 129;
Best Local Similarity 77.8%; Pred. No. 6.81e-06;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 18 VDCGSGSDPCIHRGKPF 35
: |||||:|||||
QY 15 LGGCHGSEPCNIHRGKPF 32

RESULT 14
ID US-08-461-441-8 STANDARD; PRT; 129 AA.
XX

AC xxxxxx
XX
DT
XX
DE
XX
Sequence 8, Application US/08461441
Sequence 8, Application US/08461441
Patent No. 5773002
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,441
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 129 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 129 AA; 14022 MW; 84331 CN;

Query Match 50.6%; Score 122; DB 1; Length 129;
Best Local Similarity 77.8%; Pred. No. 6.81e-06;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 18 VDCGSGSDPCIHRGKPF 35
: |||||:|||||
QY 15 LGGCHGSEPCNIHRGKPF 32

RESULT 15
ID US-08-462-831-12 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 12, Application US/08462831
Sequence 12, Application US/08462831
Patent No. 552142
GENERAL INFORMATION:
APPLICANT:

```

CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/462,831
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC APPLICATION NUMBER: US 580,655
CC FILING DATE: 11 SEPTEMBER 1990
CC APPLICATION NUMBER: US 458,642
CC FILING DATE: 13 FEBRUARY 1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 47
CC OTHER INFORMATION: /label=xaa is Thr or Ser
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 114
CC OTHER INFORMATION: /label=xaa is Asp or Asn
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 127
CC OTHER INFORMATION: /label=xaa is Ile or Leu
CC SQ SEQUENCE 129 AA; 14122 MW; 90398 CN;

Query Match 50.6%; Score 122; DB 1; Length 129;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32

Search completed: Sun Apr 2 18:10:17 2000
Job time : 4 secs.

```

WILEY

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:05:02 2000; MasPar time 2.37 Seconds
Tabular output not generated. 218.649 Million cell updates/sec

Title: >US-09-362-731-4
Description: (1-40) from US09362731.pep
Perfect Score: 278
Sequence: 1 PRYKQNTLKLATGKRGKPRYKQNTLKLATGKRGVIGIK 40

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 20.446; Variance 81.810; scale 0.250

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	110	39.6	347	1	US-08-229- Sequence 58, Applicati	5.03e-03
2	110	39.6	347	1	US-08-630- Sequence 58, Applicati	5.03e-03
3	109	39.2	25	2	US-08-484- Sequence 24, Applicati	6.23e-03
4	106	38.1	570	2	US-08-453- Sequence 7, Applicatio	1.18e-02
5	106	38.1	571	2	US-08-453- Sequence 15, Applicati	1.18e-02
6	106	38.1	571	2	US-08-453- Sequence 21, Applicati	1.18e-02
7	90	32.4	13	3	PCT-US94-1 Sequence 20, Applicati	3.39e-01
8	90	32.4	13	3	PCT-US93-0 Sequence 24, Applicati	3.39e-01
9	90	32.4	13	2	US-08-480- Sequence 24, Applicati	3.39e-01
10	90	32.4	13	2	US-08-596- Sequence 164, Applicat	3.39e-01
11	90	32.4	13	1	US-08-305- Sequence 1, Applicatio	3.39e-01
12	90	32.4	13	1	US-08-465- Sequence 22, Applicati	3.39e-01
13	90	32.4	13	3	PCT-US92-0 Sequence 18, Applicati	3.39e-01
14	90	32.4	13	1	US-07-831- Sequence 5, Applicatio	3.39e-01
15	90	32.4	13	2	US-08-596- Sequence 8, Applicatio	3.39e-01
16	90	32.4	13	2	US-08-488- Sequence 24, Applicati	3.39e-01
17	90	32.4	13	3	PCT-US95-0 Sequence 53, Applicati	3.39e-01
18	90	32.4	13	3	PCT-US95-0 Sequence 99, Applicati	3.39e-01
19	90	32.4	13	3	PCT-US95-0 Sequence 8, Applicatio	3.39e-01
20	86	30.9	13	3	PCT-US95-0 Sequence 59, Applicati	7.71e-01
21	85	30.6	13	1	US-08-619- Sequence 9, Applicatio	9.45e-01
22	85	30.6	13	2	US-08-634- Sequence 9, Applicatio	9.45e-01
23	83	29.9	13	3	PCT-US95-0 Sequence 58, Applicati	1.42e+00

24 79 28.4 13 3 PCT-US95-0 Sequence 57, Applicati 3.19e+00
25 75 27.0 348 1 US-08-229- Sequence 50, Applicati 7.08e+00
26 75 27.0 348 1 US-08-630- Sequence 50, Applicati 7.08e+00
27 70 25.2 1148 2 US-08-313- Sequence 58, Applicati 1.89e+01
28 68 24.5 572 2 US-08-453- Sequence 9, Applicatio 2.79e+01
29 67 24.1 1082 1 US-08-429- Sequence 2, Applicatio 3.38e+01
30 67 24.1 1082 1 US-08-106- Sequence 2, Applicatio 3.38e+01
31 67 24.1 1139 1 US-08-832- Sequence 2, Applicatio 3.38e+01
32 67 24.1 1139 1 US-08-832- Sequence 2, Applicatio 3.38e+01
33 66 23.7 364 3 PCT-US96-1 Sequence 2, Applicatio 4.10e+01
34 66 23.7 393 3 PCT-US96-1 Sequence 3, Applicatio 4.10e+01
35 65 23.4 446 2 US-08-244- Sequence 5, Applicatio 4.97e+01
36 65 23.4 446 2 US-08-244- Sequence 5, Applicatio 4.97e+01
37 65 23.4 446 2 US-08-244- Sequence 5, Applicatio 4.97e+01
38 64 23.0 977 2 US-08-673- Sequence 10, Applicati 4.97e+01
39 63 22.7 976 2 US-08-702- Sequence 8, Applicatio 6.01e+01
40 63 22.7 976 2 US-08-449- Sequence 18, Applicati 7.27e+01
41 63 22.7 976 2 US-08-449- Sequence 18, Applicati 7.27e+01
42 63 22.7 1147 2 US-08-668- Sequence 38, Applicati 7.27e+01
43 63 22.7 1147 1 US-08-131- Sequence 38, Applicati 7.27e+01
44 62 22.3 120 1 US-08-336- Sequence 17, Applicati 8.78e+01
45 62 22.3 554 3 PCT-US94-0 Sequence 7, Applicatio 8.78e+01

ALIGNMENTS

RESULT 1
ID US-08-229-781-58 STANDARD; PRT; 347 AA.
XX XXXXXX
DT
DE Sequence 58, Application US/08229781
XX
CC Sequence 58, Application US/08229781
CC Patent No. 5589174
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/229,781
CC FILING DATE: April 19, 1994
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single

CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 1; Length 347;
Best Local Similarity 63.3%; Pred. No. 5.03e-03;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKONTLKLATGMNV 120
|||::| |||||
QY 6 QNTLKLATGKRGPKYVKONTLKLATGKRGV 35

RESULT 2
ID US-08-630-918-58 STANDARD; PRT; 347 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 58, Application US/08630918
XX
XX Sequence 58, Application US/08630918
CC Patent No. 5631350
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/630,918
CC FILING DATE: April 5, 1996
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 1; Length 347;
Best Local Similarity 63.3%; Pred. No. 5.03e-03;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKONTLKLATGMNV 120
|||::| |||||
QY 6 QNTLKLATGKRGPKYVKONTLKLATGKRGV 35

RESULT 3
ID US-08-484-905-24 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 24, Application US/08484905
XX
XX Sequence 24, Application US/08484905
CC Patent No. 5976551
CC GENERAL INFORMATION:
CC APPLICANT: Mottez, Estelle
CC APPLICANT: Abastado, Jean-Pierre
CC APPLICANT: Kourilsky, Philippe
CC TITLE OF INVENTION: An Altered Major Histocompatibility
CC TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the
CC TITLE OF INVENTION: Determinant
CC NUMBER OF SEQUENCES: 127
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy Disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,905

```

C C LOCATION: 1 to 18
C FEATURE:
C NAME/KEY: mature rHA
C C LOCATION: 19 to 552
C C
C Q SEQUENCE 570 AA; 64054 MW; 1596768 CN;

Query Match 38.1%; Score 106; DB 2; Length 570;
Best Local Similarity 77.8%;
Matches 14; Conservative 3; Mismatches 1; Indels 1;

```

```
Db 326 PRYVKQNTLKLATGMRNV 343
      |:|:|:|:|:|:|:|:|:|
Qy 18 PKYVKQNTLKLATGKKGV 35
```

```

RESULT      5
ID US-08-453-848-15 STANDARD; PRT; 571 AA.
XX
XX xxxxxx
XX
XX
XX
XX
DE Sequence 15, Application US/08453848
XX
XX Sequence 15, Application US/08453848
CC Patent No. 5858368
CC GENERAL INFORMATION:
CC APPLICANT: Smith, Gale Eugene
CC APPLICANT: Volvovitz, Franklin
CC APPLICANT: Wilkinson, Bethanie Eident
CC APPLICANT: Voznesensky, Andrei I.
CC APPLICANT: Hackett, Craig Stanley
CC TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
CC TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Patrea L. Pabst
CC STREET: 2800 One Atlantic Center
CC STREET: 1201 West Peachtree Street
CC CITY: Atlanta
CC STATE: GA
CC COUNTRY: USA
CC ZIP: 30309-3450
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

```

CC APPLICATION NUMBER: US/08/453,848
CC FILING DATE: 30-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/120,607
CC FILING DATE: 13-SEPT-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MGS101CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-873-8794
CC TELEFAX: (404)-873-8795
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 571 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Influenza virus
CC INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
CC FEATURE:
CC NAME/KEY: AcNPV 61K protein signal sequence
CC LOCATION: 1 to 18
CC NAME/KEY: mature rHA
CC LOCATION: 19 to 553
CC SEQUENCE 571 AA; 64061 MW; 1600797 CN;
SQ
Query Match 38.1%; Score 106; DB 2; Length 571;
Best Local Similarity 77.8%; Pred. No. 1.18e-02;
Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 327 PRYVKQNTLKLATGMNV 344
QY 18 PRYVKQNTLKLATGKGV 35
RESULT 6
ID US-08-453-848-21 STANDARD; PRT: 571 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 21, Application US/08453848
XX Sequence 21, Application US/08453848
CC Patent No. 5858368
CC GENERAL INFORMATION:
CC APPLICANT: Smith, Gale Eugene
CC APPLICANT: Volnovitz, Franklin
CC APPLICANT: Wilkinson, Bethanie Eident
CC APPLICANT: Voznesensky, Andrei I.
CC APPLICANT: Hackett, Craig Stanway
CC TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
CC TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Patrea L. Pabst
CC STREET: 2800 One Atlantic Center
CC STREET: 1201 West Peachtree Street
CC CITY: Atlanta
CC STATE: GA
CC COUNTRY: USA
CC ZIP: 30309-3450
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/453,848
CC FILING DATE: 30-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/120,607
CC FILING DATE: 13-SEPT-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MGS101CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-873-8794
CC TELEFAX: (404)-873-8795
CC INFORMATION FOR SEQ ID NO: 21:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 571 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Influenza virus
CC INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA
CC FEATURE:
CC NAME/KEY: AcNPV 61K protein signal sequence
CC LOCATION: 1 to 18
CC NAME/KEY: mature rHA
CC LOCATION: 19 to 569
CC SEQUENCE 571 AA; 64271 MW; 1597291 CN;
SQ
Query Match 38.1%; Score 106; DB 2; Length 571;
Best Local Similarity 77.8%; Pred. No. 1.18e-02;
Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 327 PRYVKQNTLKLATGMNV 344
QY 18 PRYVKQNTLKLATGKGV 35
RESULT 7
ID PCT-US94-10257A-20 STANDARD; PRT: 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 20, Application PC/TUS9410257A
XX Sequence 20, Application PC/TUS9410257A
CC GENERAL INFORMATION:
CC APPLICANT: IMMULOGIC PHARMACEUTICAL CORPORATION
CC APPLICANT: BRIGITTE DEVAUX
CC APPLICANT: JONATHAN B. ROTHARD
CC APPLICANT: DAWN SMILER
CC TITLE OF INVENTION: USES OF MYELIN OLIGODENDROCYTE
CC TITLE OF INVENTION: GLYCOPROTEIN AND PEPTIDE PORTIONS THEREOF IN PROTOCOLS
CC TITLE OF INVENTION: RELATED TO AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 95
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: Massachusetts

CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 169:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SQ SEQUENCE 25 AA; 2721 MW; 3390 CN;

Query Match 100.0%; Score 106; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
| | | | | | | | | |
Qy 1 HEIKKVLVPGCHGS 14

RESULT 9
ID US-08-482-142-52 STANDARD; PRT; 26 AA.
XX
AC
XX
XX
DT
XX
DE

Sequence 52, Application US/08482142

Sequence 52, Application US/08482142
Patent No. 5820862

GENERAL INFORMATION:

CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60S
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid

CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SQ SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
| | | | | | | | | |
Qy 1 HEIKKVLVPGCHGS 14

RESULT 10
ID US-08-478-572-52 STANDARD; PRT; 26 AA.
XX
AC
XXXXXX
XX
DT
XX
DE

Sequence 52, Application US/08478572

Sequence 52, Application US/08478572
Patent No. 5968526

GENERAL INFORMATION:

CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60S
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SQ SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 11
ID US-08-478-572-50 STANDARD; PRT; 26 AA.

XX AC xxxxxx

XX DT

Sequence 50, Application US/08478572

Sequence 50, Application US/08478572

Patent No. 5968526

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/478,572

FILING DATE: 07-June-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/445,307

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017,605

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

SEQUENCE 26 AA; 2818 MW; 3780 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.32e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 12
ID US-08-482-142-170 STANDARD; PRT; 26 AA.
XX AC xxxxxx
XX DT

XX DE

XX XX

Sequence 170, Application US/08482142

Sequence 170, Application US/08482142

Patent No. 5820862

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,142

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/445,307

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017,605

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 170:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

SEQUENCE 26 AA; 2778 MW; 3598 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.32e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 13
ID US-08-482-142-50 STANDARD; PRT; 26 AA.
XX AC xxxxxx

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 170:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 26 AA; 2778 MW; 3598 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14

Search completed: Sun Apr 2 17:54:32 2000
Job time : 4 secs.

CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941

CC INFORMATION FOR SEQ ID NO: 52:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 26 amino acids

CC TYPE: AMINO ACID

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: N-terminal

CC SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 52.2%; Score 117; DB 3; Length 26;

Best Local Similarity 63.8%; Pred. No. 2.84e-04;

Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 3 VDVKDSANHEIKKVLPGCHGS 24

QY 10 IGITELGGHEIKKVLPGCHGS 31

RESULT 15

ID US-08-484-296-52 STANDARD; PRT; 26 AA.

AC xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 52, Application US/08484296

Sequence 52, Application US/08484296

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,296

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/445,307

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017.6US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 52.2%; Score 117; DB 9; Length 26;

Best Local Similarity 63.6%; Pred. No. 2.84e-04;

Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 3 VDVKDSANHEIKKVLPGCHGS 24

QY 10 IGITELGGHEIKKVLPGCHGS 31

Search completed: Sun Apr 2 17:50:05 2000

Job time : 14 secs.

```
CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Can
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488, 320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488, 351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446, 692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229, 275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057, 166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;
Query Match 52.7%; Score 118; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 21 QYIKANSKFIGITELGG 37
Qy 1 QYIKANSKFIGITELGG 17
RESULT 13
ID US-07-881-396A-51 STANDARD; PRT: 26 AA.
XX xxxxxx
AC
XX
DT
DE
XX
Sequence 51, Application US/07881396A
CC Sequence 51, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
```

```
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REFERENCE/DOCKET NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 26 AA; 2844 MW; 3796 CN;
Query Match 52.2%; Score 117; DB 3; Length 26;
Best Local Similarity 63.6%; Pred. No. 2.84e-04;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 3 VDVKDEANHEIKKVLVPGCHGS 24
Qy 10 IGITELGGHEIKKVLVPGCHGS 31
RESULT 14
ID US-07-881-396A-52 STANDARD; PRT: 26 AA.
XX xxxxxx
AC
XX
DT
DE
XX
Sequence 52, Application US/07881396A
CC Sequence 52, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
```


CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 52.7%; Score 118; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITELGG 21
QY 1 QYIKANSKFIGITELGG 17

RESULT 8
ID US-08-488-320A-63 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT

Sequence 63, Application US/08488320A

Sequence 63, Application US/08488320A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH

TITLE OF INVENTION: And A Helper T-cell Epitope For Treatment Of Prostate Cancer

TITLE OF INVENTION: And Induction Of Infertility

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488.320A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/488.351

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/446.692

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/229.275

FILING DATE: 14-APR-1994

APPLICATION NUMBER: 08/057.166

FILING DATE: 27-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-41460S4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)758-4800

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO: 57:

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 63:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 37 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 52.7%; Score 118; DB 9; Length 37;

Best Local Similarity 100.0%; Pred. No. 2.15e-04;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITELGG 21

QY 1 QYIKANSKFIGITELGG 17

RESULT 9

ID US-08-488-320A-57 STANDARD; PRT; 37 AA.

XX

AC xxxxxx

XX

DT

DE

XX

XX

Sequence 57, Application US/08488320A

Sequence 57, Application US/08488320A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH

TITLE OF INVENTION: And A Helper T-cell Epitope For Treatment Of Prostate Cancer

TITLE OF INVENTION: And Induction Of Infertility

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488.320A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/488.351

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/446.692

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/229.275

FILING DATE: 14-APR-1994

APPLICATION NUMBER: 08/057.166

FILING DATE: 27-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-41460S4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)758-4800

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO: 57:

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 29 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 29 AA; 3195 MW; 5252 CN;

Query Match 52.7%; Score 118; DB 12; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITELGG 19
QY 1 QYIKANSKFIGITELGG 17
|||||

RESULT 5
ID US-09-362-731-5 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 5, Application US/09362731
XX
XX Sequence 5, Application US/09362731
CC GENERAL INFORMATION:
CC APPLICANT: U.C.B. S.A.
CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
CC TITLE OF INVENTION: TREATMENT OF ALLERGY
CC FILE REFERENCE: P.OCB.09/WO
CC CURRENT APPLICATION NUMBER: US/09/362.731
CC CURRENT FILING DATE: 1999-07-29
CC NUMBER OF SEQ ID NOS: 17
CC SOFTWARE: PatentIn Ver. 2.1
CC SEQ ID NO 5
CC LENGTH: 32
CC TYPE: PRT
CC ORGANISM: Artificial Sequence
CC FEATURE:
CC OTHER INFORMATION: Description of Artificial Sequence: synthetic
CC OTHER INFORMATION: peptide
SQ SEQUENCE 32 AA; 3503 MW; 5050 CN;

Query Match 52.7%; Score 118; DB 18; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITELGG 17
QY 1 QYIKANSKFIGITELGG 17
|||||

RESULT 6
ID PCT-US94-04832A-57 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 57, Application PC/TUS9404832A
XX
XX Sequence 57, Application PC/TUS9404832A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs

CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 52.7%; Score 118; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITELGG 37
QY 1 QYIKANSKFIGITELGG 17
|||||

RESULT 7
ID PCT-US94-04832A-63 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 63, Application PC/TUS9404832A
XX
XX Sequence 63, Application PC/TUS9404832A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A

DE XX Sequence 3, Application US/09362731
CC Sequence 3, Application US/09362731
CC GENERAL INFORMATION:
CC APPLICANT: U.C.B. S.A.
CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
CC FILE OF INVENTION: TREATMENT OF ALLERGY
CC FILE REFERENCE: P.UCB.09/WO
CC CURRENT APPLICATION NUMBER: US/09/362,731
CC CURRENT FILING DATE: 1999-07-29
CC NUMBER OF SEQ ID NOS: 17
CC SOFTWARE: Patent in Ver. 2.1
CC SEQ ID NO 3
CC LENGTH: 137
CC TYPE: PRT
CC ORGANISM: Artificial Sequence
CC FEATURE:
CC OTHER INFORMATION: Description of Artificial Sequence: synthetic
CC OTHER INFORMATION: peptide
CC SEQUENCE 137 AA; 14747 MW; 93695 CN;
SQ

Query Match 58.0%; Score 130; DB 18; Length 137;
Best Local Similarity 90.5%; Pred. No. 7.11e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2 QYIKANSKFIGITELGGQYIK 22
QY 1 QYIKANSKFIGITELGGHEIK 21

RESULT 3
ID US-08-328-912B-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08328912B
XX
CC Sequence 37, Application US/08328912B
CC GENERAL INFORMATION:
CC APPLICANT: Wang, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 61
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/328,912B
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323

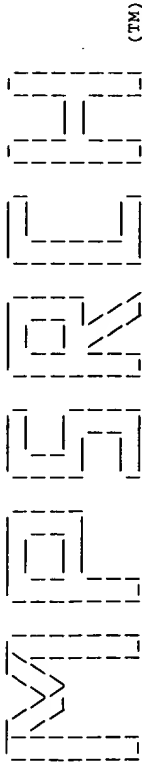
DE XX Sequence 3, Application US/08718490A
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08718490A
XX
CC Sequence 37, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323

DE XX Sequence 3, Application US/09362731
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/09362731
XX
CC Sequence 3, Application US/09362731
CC GENERAL INFORMATION:
CC APPLICANT: U.C.B. S.A.
CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
CC FILE OF INVENTION: TREATMENT OF ALLERGY
CC FILE REFERENCE: P.UCB.09/WO
CC CURRENT APPLICATION NUMBER: US/09/362,731
CC CURRENT FILING DATE: 1999-07-29
CC NUMBER OF SEQ ID NOS: 17
CC SOFTWARE: Patent in Ver. 2.1
CC SEQ ID NO 3
CC LENGTH: 137
CC TYPE: PRT
CC ORGANISM: Artificial Sequence
CC FEATURE:
CC OTHER INFORMATION: Description of Artificial Sequence: synthetic
CC OTHER INFORMATION: peptide
CC SEQUENCE 137 AA; 14747 MW; 93695 CN;
SQ

Query Match 52.7%; Score 118; DB 7; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITELGG 19
QY 1 QYIKANSKFIGITELGG 17

RESULT 4
ID US-08-718-490A-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08718490A
XX
CC Sequence 37, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:49:51 2000; MasPar time 12.10 Seconds
273.319 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-1
Description: (1-31) from US09362731.pep
Perfect Score: 224
Sequence: 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 25.291; Variance 73.379; scale 0.345

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	224	100.0	31	18	US-09-362- Sequence 1, Applicatio	3.72e-18
2	130	58.0	137	18	US-09-362- Sequence 3, Applicatio	7.11e-06
3	118	52.7	29	7	US-08-328- Sequence 37, Applicati	2.15e-04
4	118	52.7	29	12	US-08-718- Sequence 37, Applicati	2.15e-04
5	118	52.7	32	18	US-09-362- Sequence 5, Applicatio	2.15e-04
6	118	52.7	37	1	PCT-US94-0 Sequence 57, Applicati	2.15e-04
7	118	52.7	37	1	PCT-US94-0 Sequence 63, Applicati	2.15e-04
8	118	52.7	37	9	US-08-488- Sequence 63, Applicati	2.15e-04
9	118	52.7	37	9	US-08-488- Sequence 57, Applicati	2.15e-04
10	118	52.7	47	1	PCT-US94-0 Sequence 35, Applicati	2.15e-04
11	118	52.7	47	6	US-08-229- Sequence 35, Applicati	2.15e-04
12	118	52.7	47	9	US-08-488- Sequence 35, Applicati	2.15e-04
13	117	52.2	26	3	US-07-881- Sequence 51, Applicati	2.84e-04
14	117	52.2	26	3	US-07-881- Sequence 52, Applicati	2.84e-04
15	117	52.2	26	9	US-08-484- Sequence 52, Applicati	2.84e-04
16	117	52.2	26	6	US-08-227- Sequence 51, Applicati	2.84e-04
17	117	52.2	26	6	US-08-445- Sequence 52, Applicati	2.84e-04
18	117	52.2	26	6	US-08-227- Sequence 52, Applicati	2.84e-04
19	117	52.2	26	8	US-08-445- Sequence 51, Applicati	2.84e-04
20	117	52.2	26	9	US-08-484- Sequence 51, Applicati	2.84e-04

21	116	51.8	25	15	US-09-060- Sequence 37, Applicati	3.77e-04
22	116	51.8	25	15	US-09-060- Sequence 37, Applicati	3.77e-04
23	116	51.8	158	15	US-09-060- Sequence 8, Applicatio	3.77e-04
24	116	51.8	158	15	US-09-060- Sequence 8, Applicatio	3.77e-04
25	113	50.4	25	6	US-08-227- Sequence 169, Applicat	8.70e-04
26	113	50.4	25	15	US-09-060- Sequence 36, Applicati	8.70e-04
27	113	50.4	25	15	US-09-060- Sequence 36, Applicati	8.70e-04
28	113	50.4	26	6	US-08-227- Sequence 50, Applicati	8.70e-04
29	113	50.4	26	9	US-08-484- Sequence 50, Applicati	8.70e-04
30	113	50.4	26	8	US-08-445- Sequence 50, Applicati	8.70e-04
31	113	50.4	26	6	US-08-445- Sequence 170, Applicat	8.70e-04
32	113	50.4	26	6	US-08-227- Sequence 170, Applicat	8.70e-04
33	113	50.4	26	3	US-07-881- Sequence 50, Applicati	8.70e-04
34	113	50.4	26	9	US-08-484- Sequence 170, Applicat	8.70e-04
35	113	50.4	27	6	US-08-227- Sequence 171, Applicat	8.70e-04
36	113	50.4	27	8	US-08-445- Sequence 171, Applicat	8.70e-04
37	113	50.4	27	9	US-08-484- Sequence 171, Applicat	8.70e-04
38	113	50.4	35	9	US-08-460- Sequence 7, Applicatio	8.70e-04
39	113	50.4	145	9	US-08-460- Sequence 6, Applicatio	8.70e-04
40	113	50.4	146	6	US-08-227- Sequence 4, Applicatio	8.70e-04
41	113	50.4	146	9	US-08-484- Sequence 4, Applicatio	8.70e-04
42	113	50.4	146	8	US-08-445- Sequence 4, Applicatio	8.70e-04
43	113	50.4	146	3	US-07-881- Sequence 4, Applicatio	8.70e-04
44	113	50.4	158	15	US-09-060- Sequence 6, Applicatio	8.70e-04
45	113	50.4	158	15	US-09-060- Sequence 6, Applicatio	8.70e-04

ALIGNMENTS

RESULT 1 US-09-362-731-1 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
XX
DT

Sequence 1, Application US/09362731

Sequence 1, Application US/09362731

GENERAL INFORMATION:

APPLICANT: U.C.B. S.A.

TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

FILE REFERENCE: P.UCB.09/WO

CURRENT APPLICATION NUMBER: US/09/362,731

CURRENT FILING DATE: 1999-07-29

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 31

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: peptide

SEQUENCE 31 AA; 3325 MW; 5513 CN;

Query Match 100.0%; Score 224; DB 18; Length 31;

Best Local Similarity 100.0%; Pred. No. 3.72e-18;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

|||||

QY 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

RESULT 2 US-09-362-731-3 STANDARD; PRT; 137 AA.

XX

AC xxxxxx

XX

DT

XX

M P E R L F
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Apr 2 17:54:49 2000; MasPar time 10.17 Seconds
146.872 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:060 3:07 4:080 5:081 6:082 7:083 8:084A 9:084B
10:085 11:086 12:087 13:088 14:089 15:090 16:091 17:092
18:093 19:094 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 20.674; Variance 49.657; scale 0.416

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	100.0	14	US-09-362-	Sequence 2, Applicatio	1.03e-05
2	106	100.0	16	US-08-484-	Sequence 59, Applicati	1.03e-05
3	106	100.0	16	US-08-227-	Sequence 59, Applicati	1.03e-05
4	106	100.0	16	US-08-445-	Sequence 59, Applicati	1.03e-05
5	106	100.0	16	US-07-881-	Sequence 59, Applicati	1.03e-05
6	106	100.0	25	US-07-881-	Sequence 42, Applicati	1.03e-05
7	106	100.0	25	US-08-227-	Sequence 169, Applicati	1.03e-05
8	106	100.0	25	US-08-445-	Sequence 42, Applicati	1.03e-05
9	106	100.0	25	US-08-445-	Sequence 169, Applicati	1.03e-05
10	106	100.0	25	US-08-484-	Sequence 169, Applicati	1.03e-05
11	106	100.0	25	US-08-227-	Sequence 42, Applicati	1.03e-05
12	106	100.0	25	US-08-484-	Sequence 42, Applicati	1.03e-05
13	106	100.0	26	US-08-227-	Sequence 51, Applicati	1.03e-05
14	106	100.0	26	US-07-881-	Sequence 50, Applicati	1.03e-05
15	106	100.0	26	US-07-881-	Sequence 51, Applicati	1.03e-05
16	106	100.0	26	US-08-445-	Sequence 170, Applicat	1.03e-05
17	106	100.0	26	US-08-484-	Sequence 51, Applicati	1.03e-05
18	106	100.0	26	US-08-227-	Sequence 170, Applicat	1.03e-05
19	106	100.0	26	US-08-445-	Sequence 50, Applicati	1.03e-05
20	106	100.0	26	US-08-484-	Sequence 52, Applicati	1.03e-05

21	106	100.0	26	US-08-227-	Sequence 52, Applicati	1.03e-05
22	106	100.0	26	US-07-881-	Sequence 52, Applicati	1.03e-05
23	106	100.0	26	US-08-484-	Sequence 50, Applicati	1.03e-05
24	106	100.0	26	US-08-445-	Sequence 51, Applicati	1.03e-05
25	106	100.0	26	US-08-227-	Sequence 50, Applicati	1.03e-05
26	106	100.0	26	US-08-484-	Sequence 170, Applicat	1.03e-05
27	106	100.0	26	US-08-445-	Sequence 52, Applicati	1.03e-05
28	106	100.0	27	US-08-227-	Sequence 171, Applicat	1.03e-05
29	106	100.0	27	US-08-484-	Sequence 171, Applicat	1.03e-05
30	106	100.0	27	US-08-445-	Sequence 171, Applicat	1.03e-05
31	106	100.0	31	US-09-362-	Sequence 1, Applicatio	1.03e-05
32	106	100.0	35	US-08-460-	Sequence 7, Applicatio	1.03e-05
33	106	100.0	145	US-08-460-	Sequence 6, Applicatio	1.03e-05
34	106	100.0	146	US-08-445-	Sequence 4, Applicatio	1.03e-05
35	106	100.0	146	US-08-227-	Sequence 4, Applicatio	1.03e-05
36	106	100.0	146	US-07-881-	Sequence 4, Applicatio	1.03e-05
37	106	100.0	146	US-08-484-	Sequence 4, Applicatio	1.03e-05
38	95	89.6	61	US-08-445-	Sequence 162, Applicat	4.47e-04
39	95	89.6	61	US-08-227-	Sequence 162, Applicat	4.47e-04
40	95	89.6	84	US-08-227-	Sequence 161, Applicat	4.47e-04
41	95	89.6	129	US-08-227-	Sequence 159, Applicat	4.47e-04
42	95	89.6	129	US-08-227-	Sequence 157, Applicat	4.47e-04
43	95	89.6	129	US-08-445-	Sequence 159, Applicat	4.47e-04
44	95	89.6	130	US-08-445-	Sequence 158, Applicat	4.47e-04
45	95	89.6	130	US-08-227-	Sequence 158, Applicat	4.47e-04

ALIGNMENTS

RESULT 1
ID US-09-362-731-2 STANDARD; PRT; 14 AA.

XX xxxxxx

DT

XX Sequence 2, Application US/09362731

DE Sequence 2, Application US/09362731

CC Sequence 2, Application US/09362731

CC GENERAL INFORMATION:

CC APPLICANT: U.C.B. S.A.

CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

CC FILE REFERENCE: P.UCB.09/WO

CC CURRENT APPLICATION NUMBER: US/09/362,731

CC CURRENT FILING DATE: 1999-07-29

CC NUMBER OF SEQ ID NOS: 17

CC SOFTWARE: PatentIn Ver. 2.1

CC SEQ ID NO 2

CC LENGTH: 14

CC TYPE: PRT

CC ORGANISM: Artificial Sequence

CC FEATURE:

CC OTHER INFORMATION: Description of Artificial Sequence: synthetic

CC OTHER INFORMATION: peptide

SQ SEQUENCE 14 AA; 1504 MW; 1224 CN;

Query Match 100.0%; Score 106; DB 18; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14

QY 1 HEIKKVLVPGCHGS 14

RESULT 2
ID US-08-484-296-59 STANDARD; PRT; 16 AA.

XX xxxxxx

DT

XX

DE Sequence 59, Application US/08484296
XX
CC Sequence 59, Application US/08484296
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 59:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 16 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 16 AA; 1730 MW; 1569 CN;
SQ
Query Match 100.0%; Score 106; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14
RESULT 3
ID US-08-227-772-59 STANDARD; PRT; 16 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
CC Sequence 59, Application US/08227772
CC Sequence 59, Application US/08227772
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia

CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 197
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 59:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 16 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 16 AA; 1730 MW; 1569 CN;
SQ
Query Match 100.0%; Score 106; DB 6; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14
RESULT 4
ID US-08-445-307A-59 STANDARD; PRT; 16 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
CC Sequence 59, Application US/08445307A
CC Sequence 59, Application US/08445307A
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM

CC	REFERENCE/DOCKET NUMBER:	IMI-012/IPC-017C	
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE:	(617) 227-7400	
CC	TELEFAX:	(617) 227-5941	
CC	INFORMATION FOR SEQ ID NO:	59:	
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH:	16 amino acids	
CC	TYPE:	AMINO ACID	
CC	TOPOLOGY:	linear	
CC	MOLECULE TYPE:	peptide	
CC	FRAGMENT TYPE:	N-terminal	
CC	SEQUENCE	16 AA; 1730 MW; 1569 CN;	
CC	Query Match	100.0%; Score 106; DB 3; Length 16;	
CC	Best Local Similarity	100.0%; Pred. No. 1.03e-05;	
CC	Matches	14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
CC	Db	1 HEIKKVLVPGCHGS 14	
CC	Qy	1 HEIKKVLVPGCHGS 14	
CC	RESULT	6	
CC	ID	US-07-881-396A-42	
CC	XX	STANDARD; PRT; 25 AA.	
CC	XX	xxxxxx	
CC	DT		
CC	XX		
CC	DE	Sequence 42, Application US/07881396A	
CC	XX		
CC	CC	Sequence 42, Application US/07881396A	
CC	CC	GENERAL INFORMATION:	
CC	CC	APPLICANT: Kuo, Mei-chang	
CC	CC	APPLICANT: Garman, Richard	
CC	CC	APPLICANT: Greenstein, Julia	
CC	CC	TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS	
CC	CC	TYPE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)	
CC	CC	NUMBER OF SEQUENCES: 119	
CC	CC	CORRESPONDENCE ADDRESS:	
CC	CC	ADDRESSEE: LAHIVE & COCKFIELD	
CC	CC	STREET: 60 STATE STREET SUITE 510	
CC	CC	CITY: BOSTON	
CC	CC	STATE: MA	
CC	CC	COUNTRY: USA	
CC	CC	ZIP: 02109	
CC	CC	COMPUTER READABLE FORM:	
CC	CC	MEDIUM TYPE: Floppy disk	
CC	CC	COMPUTER: IBM PC compatible	
CC	CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	CC	SOFTWARE: ASCII TEXT	
CC	CC	CURRENT APPLICATION DATA:	
CC	CC	APPLICATION NUMBER: US/07/881.396A	
CC	CC	FILING DATE: 19920508	
CC	CC	CLASSIFICATION: 424	
CC	CC	ATTORNEY/AGENT INFORMATION:	
CC	CC	NAME: MANDRAGOURAS, AMY E.	
CC	CC	REGISTRATION NUMBER: P36,207	
CC	CC	REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C	
CC	CC	TELECOMMUNICATION INFORMATION:	
CC	CC	TELEPHONE: (617) 227-7400	
CC	CC	TELEFAX: (617) 227-5941	
CC	CC	INFORMATION FOR SEQ ID NO: 42:	
CC	CC	SEQUENCE CHARACTERISTICS:	
CC	CC	LENGTH: 25 amino acids	
CC	CC	TYPE: AMINO ACID	
CC	CC	TOPOLOGY: linear	
CC	CC	MOLECULE TYPE: peptide	
CC	CC	FRAGMENT TYPE: N-terminal	
CC	CC	SEQUENCE	25 AA; 2782 MW; 3408 CN;
CC	Query Match	100.0%; Score 106; DB 3; Length 25;	
CC	Best Local Similarity	100.0%; Pred. No. 1.03e-05;	

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 169:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 25 AA; 2721 MW; 3390 CN;
SQ
Query Match 100.0%; Score 106; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
RESULT 10
ID US-08-484-296-169 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
DT
XX
DE Sequence 169, Application US/08484296
CC
CC Sequence 169, Application US/08484296
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET

CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 169:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 25 AA; 2721 MW; 3390 CN;
SQ
Query Match 100.0%; Score 106; DB 9; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
RESULT 11
ID US-08-227-772-42 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
DT
XX
DE Sequence 42, Application US/08227772
CC
CC Sequence 42, Application US/08227772
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 197
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
SQ

Query Match 100.0%; Score 106; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14

RESULT 12
ID US-08-484-296-42 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
DT
XX

Sequence 42, Application US/08484296
Sequence 42, Application US/08484296
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
SQ

Query Match 100.0%; Score 106; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14

RESULT 13
ID US-08-227-772-51 STANDARD; PRT: 26 AA.
XX
AC xxxxxx
XX
DT
DT
XX

Sequence 51, Application US/08227772
Sequence 51, Application US/08227772
GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,772
FILING DATE: 14-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/881,396
FILING DATE: 08-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
SEQUENCE 26 AA; 2844 MW; 3796 CN;
SQ

Query Match 100.0%; Score 106; DB 6; Length 26;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
 QY 1 HEIKKVLVPGCHGS 14
 |||||

RESULT 14
 ID US-07-881-396A-50 STANDARD; PRT; 26 AA.
 XX AC xxxxxx
 XX DT

Sequence 50, Application US/07881396A

Sequence 50, Application US/07881396A
 GENERAL INFORMATION:
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 119
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 STATE STREET SUITE 510
 CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207
 REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: N-terminal
 SEQUENCE 26 AA; 2818 MW; 3780 CN;

Query Match 100.0%; Score 106; DB 3; Length 26;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
 QY 1 HEIKKVLVPGCHGS 14
 |||||

RESULT 15
 ID US-07-881-396A-51 STANDARD; PRT; 26 AA.
 XX AC xxxxxx
 XX DT

XX

DT
 XX
 DE
 XX

Sequence 51, Application US/07881396A

Sequence 51, Application US/07881396A
 GENERAL INFORMATION:
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 119
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 STATE STREET SUITE 510
 CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207
 REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 51:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: N-terminal
 SEQUENCE 26 AA; 2844 MW; 3796 CN;

Query Match 100.0%; Score 106; DB 3; Length 26;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
 QY 1 HEIKKVLVPGCHGS 14
 |||||

Search completed: Sun Apr 2 17:55:01 2000
 Job time : 12 secs.

MUSEA
***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Apr 2 18:05:25 2000; MapPar time 13.09 Seconds
325.920 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-4
Description: (1-40) from US09362731.pep
Perfect Score: 278
Sequence: 1 PKYVQNTLKLATGKGPYVQNTLKLATGKKGVIIGIK 40

Scoring table: PAM 150
Gap 11

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 24.786; Variance 85.660; scale 0.289

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB	ID	Description
1	278	100.0	40	18	US-09-362- Sequence 4, Applicatio
2	110	39.6	347	8	US-08-443- Sequence 58, Applicati
3	110	39.6	347	15	US-09-004- Sequence 58, Applicati
4	110	39.6	347	4	US-08-054- Sequence 58, Applicati
5	109	39.2	23	23	US-08-944- Sequence 3, Applicatio
6	109	39.2	23	10	US-08-537- Sequence 25, Applicatio
7	109	39.2	23	14	US-08-944- Sequence 3, Applicatio
8	109	39.2	24	23	US-08-944- Sequence 11, Applicati
9	109	39.2	24	14	US-08-944- Sequence 11, Applicati
10	109	39.2	25	15	US-09-082- Sequence 1063, Applic
11	109	39.2	25	11	US-08-651- Sequence 15, Applicati
12	109	39.2	25	7	US-08-364- Sequence 24, Applicati
13	109	39.2	25	7	US-08-370- Sequence 24, Applicati
14	109	39.2	25	17	US-09-248- Sequence 85, Applicati
15	109	39.2	25	15	US-09-082- Sequence 1063, Applic
16	109	39.2	25	1	PCT-US99-1 Sequence 1063, Applic
17	109	39.2	25	1	PCT-US99-0 Sequence 85, Applicati
18	109	39.2	25	9	US-08-487- Sequence 24, Applicati
19	109	39.2	25	9	US-08-481- Sequence 24, Applicati
20	109	39.2	25	3	US-07-801- Sequence 24, Applicati

21	109	39.2	25	9	US-08-481- Sequence 24, Applicati	3.00e-02
22	109	39.2	25	11	US-08-651- Sequence 15, Applicati	3.00e-02
23	109	39.2	29	9	US-08-477- Sequence 14, Applicati	3.00e-02
24	106	38.1	23	14	US-08-944- Sequence 26, Applicati	6.05e-02
25	106	38.1	570	16	US-09-169- Sequence 7, Applicati	6.05e-02
26	106	38.1	571	16	US-09-169- Sequence 15, Applicati	6.05e-02
27	106	38.1	571	16	US-09-169- Sequence 21, Applicati	6.05e-02
28	102	36.7	16	5	US-08-121- Sequence 7, Applicati	1.53e-01
29	102	36.7	19	5	US-08-121- Sequence 1, Applicati	1.53e-01
30	101	36.3	23	10	US-08-537- Sequence 26, Applicati	1.93e-01
31	97	34.9	14	23	US-08-944- Sequence 15, Applicati	4.83e-01
32	97	34.9	14	14	US-08-944- Sequence 15, Applicati	4.83e-01
33	97	34.9	14	16	US-09-125- Sequence 61, Applicati	4.83e-01
34	97	34.9	15	16	US-09-125- Sequence 60, Applicati	4.83e-01
35	94	33.8	15	5	US-08-121- Sequence 8, Applicati	9.56e-01
36	92	33.1	15	18	US-09-359- Sequence 4, Applicati	1.50e+00
37	92	33.1	15	1	PCT-US99-1 Sequence 44, Applicati	1.50e+00
38	92	33.1	16	23	US-08-944- Sequence 6, Applicati	1.50e+00
39	92	33.1	16	14	US-08-944- Sequence 6, Applicati	1.50e+00
40	90	32.4	13	1	PCT-US99-2 Sequence 14, Applicati	2.35e+00
41	90	32.4	13	7	US-08-350- Sequence 1, Applicati	2.35e+00
42	90	32.4	13	12	US-08-731- Sequence 53, Applicati	2.35e+00
43	90	32.4	13	6	US-08-222- Sequence 1, Applicati	2.35e+00
44	90	32.4	13	9	US-08-485- Sequence 1, Applicati	2.35e+00
45	90	32.4	13	9	US-08-487- Sequence 24, Applicati	2.35e+00

ALIGNMENTS

RESULT 1
ID US-09-362-731-4 STANDARD; PRT; 40 AA.
XX
AC xxxxxx
XX
DT
XX

Sequence 4, Application US/09362731

Sequence 4, Application US/09362731

GENERAL INFORMATION:

APPLICANT: U.C.B. S.A.

TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

FILE OF INVENTION: TREATMENT OF ALLERGY

FILE REFERENCE: P.DCB.09/WO

CURRENT APPLICATION NUMBER: US/09/362,731

CURRENT FILING DATE: 1999-07-29

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 40

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: peptide

SEQUENCE 40 AA; 4354 MW; 9180 CN;

Query Match 100.0%; Score 278; DB 18; Length 40;

Best Local Similarity 100.0%; Pred. No. 4.93e-21;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 PKYVQNTLKLATGKGPYVQNTLKLATGKKGVIIGIK 40

|||||

1 PKYVQNTLKLATGKGPYVQNTLKLATGKKGVIIGIK 40

|||||

RESULT 2

ID US-08-443-862-58 STANDARD; PRT; 347 AA.

XX

AC xxxxxx

XX

DT

XX

DE Sequence 58, Application US/08443862
XX
CC Sequence 58, Application US/08443862
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/443,862
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:

CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;
Query Match 39.6%; Score 110; DB 8; Length 347;
Best Local Similarity 63.3%; Pred No. 2,37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 92 QNVKITYGA-CPKYVKONTLKLATGMNV 120
II :: I |||||IIIIII:: I
QY 6 QNTLKLATGKGKPKYVKONTLKLATGKGKGV 35
RESULT 3
ID US-09-004-422-58 STANDARD: PRT; 347 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 58, Application US/09004422
XX
CC Sequence 58, Application US/09004422
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
CC STREET: 2033 K Street, N.W., #800
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20006
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/004,422
CC FILING DATE: January 8, 1998
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/443,862
CC FILING DATE: May 22, 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-721-8200
CC TELEFAX: 202-721-8250
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide

CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC PUBLICATION DATE:
CC FILING DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
CC SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 15; Length 347;
Best Local Similarity 63.3%; Pred. No. 2.37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKQNTKLATGMNV 120
QY 6 QNTLKLATGKGGPKYVKQNTKLATGKGV 35

RESULT 4
ID US-08-054-016B-58 STANDARD; PRT; 347 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 58, Application US/08054016B
XX Sequence 58, Application US/08054016B
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC City: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS

CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/054,016B
CC FILING DATE: April 29, 1993
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC PUBLICATION DATE:
CC FILING DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
CC SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 4; Length 347;
Best Local Similarity 63.3%; Pred. No. 2.37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKQNTKLATGMNV 120
QY 6 QNTLKLATGKGGPKYVKQNTKLATGKGV 35

RESULT 5
ID US-08-944-147-3 STANDARD; PRT; 23 AA.

```
XX      xxxxxx
AC      Sequence 3, Application US/08944147A
DT      Sequence 3, Application US/08944147A
XX      GENERAL INFORMATION:
XX      APPLICANT: JACKSON, DAVID C.
CC      APPLICANT: O'BRIEN-SIMPSON, NEIL M.
CC      APPLICANT: BROWN, LORENA E.
CC      APPLICANT: EDE, NICHOLAS J.
CC      APPLICANT: BRANDT, EVELYN R.
CC      APPLICANT: GOOD, MICHAEL F.
CC      TITLE OF INVENTION: POLYMERS INCORPORATING PEPTIDES
CC      FILE REFERENCE: FBRC:006
CC      CURRENT APPLICATION NUMBER: US/08/944,147A
CC      CURRENT FILING DATE: 1997-10-06
CC      EARLIER APPLICATION NUMBER: P05071
CC      EARLIER FILING DATE: 1997-02-11
CC      NUMBER OF SEQ ID NOS: 28
CC      SOFTWARE: PatentIn Ver. 2.1
CC      SEQ ID NO 3
CC      LENGTH: 23
CC      TYPE: PRT
CC      ORGANISM: Artificial Sequence
CC      FEATURE:
CC      OTHER INFORMATION: Description of Artificial Sequence: Synthetic
CC      OTHER INFORMATION: Peptide
CC      SEQUENCE 23 AA: 2645 MW; 2952 CN;

Query Match      39.2%; Score 109; DB 23; Length 23;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      1 PKYVKQNTLKLATGMNV 18
QY      18 PKYVKQNTLKLATGKGV 35

RESULT 6
ID      US-08-537-928-25      STANDARD;      PRT;      23 AA.
XX      xxxxxx
AC      Sequence 25, Application US/08537928
DT      Sequence 25, Application US/08537928
XX      GENERAL INFORMATION:
CC      APPLICANT: Keith, Rose
CC      APPLICANT: Offord, Robin E
CC      TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
CC      FILE REFERENCE: GFEN-001/03US
CC      CURRENT APPLICATION NUMBER: US/08/537,928
CC      CURRENT FILING DATE: 1996-01-05
CC      EARLIER APPLICATION NUMBER: PCT/IB94/00093
CC      EARLIER FILING DATE: 1994-05-05
CC      EARLIER APPLICATION NUMBER: 08/105,904
CC      EARLIER FILING DATE: 1993-08-31
CC      EARLIER APPLICATION NUMBER: 08/114,877
CC      EARLIER FILING DATE: 1993-08-31
CC      NUMBER OF SEQ ID NOS: 30
CC      SOFTWARE: PatentIn Ver. 2.1
CC      SEQ ID NO 25
CC      LENGTH: 23
CC      TYPE: PRT
CC      ORGANISM: Artificial Sequence
CC      FEATURE:
CC      OTHER INFORMATION: Description of Artificial Sequence: Synthetic
CC      SEQUENCE 23 AA: 2645 MW; 2952 CN;
```

```
Query Match      39.2%; Score 109; DB 10; Length 23;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      1 PKYVKQNTLKLATGMNV 18
QY      18 PKYVKQNTLKLATGKGV 35

RESULT 7
ID      US-08-944-147-3      STANDARD;      PRT;      23 AA.
XX      xxxxxx
AC      Sequence 3, Application US/08944147
DT      Sequence 3, Application US/08944147
XX      GENERAL INFORMATION:
CC      APPLICANT: Jackson, David C.
CC      APPLICANT: O'Brien-Simpson, Neil M.
CC      APPLICANT: Brown, Lorena E.
CC      APPLICANT: Ede, Nicholas J.
CC      APPLICANT: Brandy, Evelyn R.
CC      APPLICANT: Good, Michael F.
CC      APPLICANT: Zeng, Weiguang
CC      TITLE OF INVENTION: POLYMERS INCORPORATION PEPTIDES
CC      NUMBER OF SEQUENCES: 29
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Arnold, White & Durkee
CC      STREET: P.O. Box 4433
CC      CITY: Houston
CC      STATE: TX
CC      COUNTRY: USA
CC      ZIP: 77210-4433
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/944,147
CC      FILING DATE: 06-OCT-1997
CC      CLASSIFICATION: 526
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: McMillian, Nabeela R.
CC      REGISTRATION NUMBER: P-43,363
CC      REFERENCE/DOCKET NUMBER: FBRC:006
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (512)418-3000
CC      TELEFAX: (512)474-7544
CC      INFORMATION FOR SEQ ID NO: 3:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 23 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS:
CC      TOPOLOGY: linear
CC      SEQUENCE 23 AA: 2645 MW; 2952 CN;

Query Match      39.2%; Score 109; DB 14; Length 23;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      1 PKYVKQNTLKLATGMNV 18
QY      18 PKYVKQNTLKLATGKGV 35

RESULT 8
ID      US-08-944-147-11      STANDARD;      PRT;      24 AA.
XX
```


Db 2 PKYVKQNTLKLATGMRNV 19
||| ||| ||| ||| ||| : |

Search completed: Sun Apr 2 18:
Job time : 17 secs.

W P E R L A (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Apr 2 18:10:34 2000; MasPar time 12.24 Seconds
278.714 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QVIKANSKFIGITELGGCHGSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1:PG 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 24.794; Variance 65.012; scale 0.381

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	241	100.0	32	18	US-09-362-Sequence 5, Applicatio	3.35e-23
2	221	91.7	137	18	US-09-362-Sequence 3, Applicatio	3.33e-20
3	122	50.6	25	8	US-08-445-Sequence 104, Applicat	5.99e-06
4	122	50.6	25	9	US-08-484-Sequence 42, Applicati	5.99e-06
5	122	50.6	25	3	US-07-881-Sequence 104, Applicat	5.99e-06
6	122	50.6	25	3	US-07-881-Sequence 42, Applicati	5.99e-06
7	122	50.6	25	8	US-08-445-Sequence 42, Applicati	5.99e-06
8	122	50.6	25	6	US-08-227-Sequence 42, Applicati	5.99e-06
9	122	50.6	25	9	US-08-484-Sequence 104, Applicat	5.99e-06
10	122	50.6	35	9	US-08-460-Sequence 7, Applicatio	5.99e-06
11	122	50.6	61	9	US-08-484-Sequence 162, Applicat	5.99e-06
12	122	50.6	61	6	US-08-227-Sequence 162, Applicat	5.99e-06
13	122	50.6	61	8	US-08-445-Sequence 162, Applicat	5.99e-06
14	122	50.6	84	6	US-08-227-Sequence 161, Applicat	5.99e-06
15	122	50.6	84	8	US-08-445-Sequence 161, Applicat	5.99e-06
16	122	50.6	84	9	US-08-484-Sequence 161, Applicat	5.99e-06
17	122	50.6	92	6	US-08-227-Sequence 160, Applicat	5.99e-06
18	122	50.6	92	8	US-08-445-Sequence 160, Applicat	5.99e-06
19	122	50.6	92	9	US-08-484-Sequence 160, Applicat	5.99e-06
20	122	50.6	129	14	US-08-930-Sequence 8, Applicatio	5.99e-06

21	122	50.6	129	6	US-08-227-Sequence 157, Applicat	5.99e-06
22	122	50.6	129	9	US-08-484-Sequence 157, Applicat	5.99e-06
23	122	50.6	129	9	US-08-484-Sequence 159, Applicat	5.99e-06
24	122	50.6	129	8	US-08-445-Sequence 159, Applicat	5.99e-06
25	122	50.6	129	14	US-08-930-Sequence 6, Applicatio	5.99e-06
26	122	50.6	129	14	US-08-930-Sequence 24, Applicati	5.99e-06
27	122	50.6	129	14	US-08-930-Sequence 18, Applicati	5.99e-06
28	122	50.6	129	14	US-08-930-Sequence 10, Applicati	5.99e-06
29	122	50.6	129	14	US-08-930-Sequence 2, Applicatio	5.99e-06
30	122	50.6	129	14	US-08-930-Sequence 12, Applicati	5.99e-06
31	122	50.6	129	14	US-08-930-Sequence 14, Applicati	5.99e-06
32	122	50.6	129	6	US-08-227-Sequence 8, Applicatio	5.99e-06
33	122	50.6	129	3	US-07-881-Sequence 8, Applicatio	5.99e-06
34	122	50.6	129	8	US-08-445-Sequence 8, Applicatio	5.99e-06
35	122	50.6	129	8	US-08-445-Sequence 157, Applicat	5.99e-06
36	122	50.6	129	9	US-08-484-Sequence 8, Applicatio	5.99e-06
37	122	50.6	130	6	US-08-227-Sequence 158, Applicat	5.99e-06
38	122	50.6	130	9	US-08-484-Sequence 158, Applicat	5.99e-06
39	122	50.6	130	8	US-08-445-Sequence 158, Applicat	5.99e-06
40	122	50.6	142	9	US-08-483-Sequence 2, Applicatio	5.99e-06
41	122	50.6	142	9	US-08-483-Sequence 4, Applicatio	5.99e-06
42	122	50.6	145	9	US-08-460-Sequence 6, Applicatio	5.99e-06
43	122	50.6	146	9	US-08-484-Sequence 4, Applicatio	5.99e-06
44	122	50.6	146	8	US-08-445-Sequence 4, Applicatio	5.99e-06
45	122	50.6	146	3	US-07-881-Sequence 4, Applicatio	5.99e-06

ALIGNMENTS

RESULT 1
ID US-09-362-731-5 STANDARD; PRT: 32 AA.

XX xxxxxx

AC

DT

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

DE Sequence 3, Application US/09362731
XX
CC Sequence 3, Application US/09362731
CC GENERAL INFORMATION:
CC APPLICANT: U.C.B. S.A.
CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
CC FILE OF INVENTION: TREATMENT OF ALLERGY
CC FILE REFERENCE: P.UCB.09/MO
CC CURRENT APPLICATION NUMBER: US/09/362,731
CC CURRENT FILING DATE: 1999-07-29
CC NUMBER OF SEQ ID NOS: 17
CC SOFTWARE: Patentin ver. 2.1
CC SEQ ID NO 3
CC LENGTH: 137
CC TYPE: PRT
CC ORGANISM: Artificial Sequence
CC FEATURE:
CC OTHER INFORMATION: Description of Artificial Sequence: synthetic
CC OTHER INFORMATION: peptide
CC OTHER INFORMATION: peptide
SQ SEQUENCE 137 AA: 14747 MW: 93695 CN;

Query Match 91.7%; Score 221; DB 18; Length 137;
Best Local Similarity 90.6%; Pred. No. 3.33e-20;
Matches 29; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 19 QYIKANSKEFIGITELSSCHGSEPCIIHRGKPF 50
QY 1 QYIKANSKEFIGITELGCGHSEPCNIHRGKPF 32
|||||

RESULT 3
ID US-08-445-307A-104 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
DT
DT
XX
DE Sequence 104, Application US/08445307A
XX

Sequence 104, Application US/08445307A
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,307A
FILING DATE: 07 June 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017.605
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
TYPE: amino acid
LENGTH: 25 amino acids
TOPOLOGY: linear

CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 104:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA: 2781 MW: 3272 CN;
SQ

Query Match 50.6%; Score 122; DB 8; Length 25;
Best Local Similarity 77.8%; Pred. No. 5.99e-06;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 8 VDGCHGSDPCIIHRGKPF 25
QY 15 LGCGHSEPCNIHRGKPF 32
|||||

RESULT 4
ID US-08-484-296-42 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
DT
DT
XX
DE Sequence 42, Application US/08484296
XX

Sequence 42, Application US/08484296
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 510 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017.605
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
TYPE: amino acid
LENGTH: 25 amino acids
TOPOLOGY: linear

```

XX      xxxxxx
XX      AC
XX      DT
XX      DE
XX      SEQUENCE 42, Application US/07881396A
XX
XX      SEQUENCE 42, Application US/07881396A
XX      GENERAL INFORMATION:
XX      APPLICANT: Kuo, Mei-chang
XX      APPLICANT: Garman, Richard
XX      APPLICANT: Greenstein, Julia
XX      TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
XX      TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
XX      NUMBER OF SEQUENCES: 119
XX      CORRESPONDENCE ADDRESS:
XX      ADDRESSEE: LAHIVE & COCKFIELD
XX      STREET: 60 STATE STREET SUITE 510
XX      CITY: BOSTON
XX      STATE: MA
XX      COUNTRY: USA
XX      ZIP: 02109
XX      COMPUTER READABLE FORM:
XX      MEDIUM TYPE: Floppy disk
XX      COMPUTER: IBM PC compatible
XX      OPERATING SYSTEM: PC-DOS/MS-DOS
XX      SOFTWARE: ASCII TEXT
XX      CURRENT APPLICATION DATA:
XX      APPLICATION NUMBER: US/07/881,396A
XX      FILING DATE: 19920508
XX      CLASSIFICATION: 424
XX      ATTORNEY/AGENT INFORMATION:
XX      NAME: MANDRAGOURAS, AMY E.
XX      REGISTRATION NUMBER: P36,207
XX      REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
XX      TELECOMMUNICATION INFORMATION:
XX      TELEPHONE: (617) 227-7400
XX      TELEFAX: (617) 227-5941
XX      INFORMATION FOR SEQ ID NO: 42:
XX      SEQUENCE CHARACTERISTICS:
XX      LENGTH: 25 amino acids
XX      TYPE: AMINO ACID
XX      TOPOLOGY: linear
XX      MOLECULE TYPE: peptide
XX      FRAGMENT TYPE: N-terminal
XX      SEQUENCE 25 AA; 2782 MW; 3408 CN;
XX
XX      Query Match 50.6%; Score 122; DB 3; Length 25;
XX      Best Local Similarity 93.8%; Pred. No. 5.99e-06;
XX      Matches 15; Conservative 0; Mismatches 1; Indels 0;
XX
XX      Db 10 GCHGSEPCIHRGKPF 25
XX      ||||||| |||||
XX      Qy 17 GCHGSEPCIHRGKPF 32
XX
XX      RESULT 7
XX      ID US-08-445-307A-42 STANDARD; PRT; 25 AA.
XX      AC xxxxxx
XX      DT
XX      DE SEQUENCE 42, Application US/08445307A
XX
XX      SEQUENCE 42, Application US/08445307A
XX      GENERAL INFORMATION:
XX      APPLICANT: Garman, Richard
XX      APPLICANT: Greenstein, Julia
XX      APPLICANT: Kuo, Mei-chang
XX      APPLICANT: Rogers, Bruce
XX      APPLICANT: Franzen, Henry
XX      APPLICANT: Chen, Xian

```

CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
SQ
Query Match 50.6%; Score 122; DB 8; Length 25;
Best Local Similarity 93.8%; Pred. No. 5,99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 10 GCHGSEPCIIHRGKPF 25
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 8
ID US-08-227-772-42 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 42, Application US/08227772
Sequence 42, Application US/08227772
GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
SQ
Query Match 50.6%; Score 122; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 5,99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 10 GCHGSEPCIIHRGKPF 25
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 9
ID US-08-484-296-104 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 104, Application US/08484296
Sequence 104, Application US/08484296
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

CC	INFORMATION FOR SEQ ID NO: 7:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 35 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: peptide	
CC	FRAGMENT TYPE: internal	
SQ	SEQUENCE 35 AA; 3870 MW; 6469 CN;	
	Query Match 50.6%; Score 122; DB 9; Length 35;	
	Best Local Similarity 93.8%; Pred. No. 5.99e-06;	
	Matches 15; Conservative 0; Mismatches 1; Indels 0;	
Db	20 GCHGSEPCIIHRGKPF 35	
QY	17 GCHGSEPCNIHRGKPF 32	
RESULT	11	
ID	US-08-484-296-162	STANDARD; PRT; 61 AA.
XX	AC	
XX	xxxxxx	
DT		
XX		
DE		
XX		
CC	Sequence 162, Application US/08484296	
CC	Sequence 162, Application US/08484296	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Garman, Richard	
CC	APPLICANT: Greenstein, Julia	
CC	APPLICANT: Kuo, Mei-chang	
CC	APPLICANT: Rogers, Bruce	
CC	APPLICANT: Franzen, Henry	
CC	APPLICANT: Chen, Xian	
CC	APPLICANT: Evans, Sean	
CC	APPLICANT: Shaked, Ze'ev	
CC	TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS	
CC	TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)	
CC	NUMBER OF SEQUENCES: 207	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION	
CC	STREET: 610 LINCOLN STREET	
CC	CITY: WALTHAM	
CC	STATE: MA	
CC	COUNTRY: USA	
CC	ZIP: 02154	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: ASCII TEXT	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/484,296	
CC	FILING DATE:	
CC	CLASSIFICATION: 435	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: 08/445,307	
CC	FILING DATE: 07 June 1995	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: CRAIG, ANNE I.	
CC	REGISTRATION NUMBER: 32,976	
CC	REFERENCE/DOCKET NUMBER: 017.6US	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (617) 466-6000	
CC	TELEFAX: (617) 466-6040	
CC	INFORMATION FOR SEQ ID NO: 162:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 61 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: peptide	
CC	FRAGMENT TYPE: internal	

SQ SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 50.6%; Score 122; DB 9; Length 61;
 Best Local Similarity 93.8%; Pred. No. 5.99e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
 ||||||| |||||||
 QY 17 GCHGSEPCNIHRGKPF 32

RESULT 12

ID US-08-227-772-162 STANDARD; PRT; 61 AA.
 XX
 AC
 DT
 XX
 XX
 DE
 XX
 XX

Sequence 162, Application US/08227772

Sequence 162, Application US/08227772
 GENERAL INFORMATION:
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Rogers, Bruce
 APPLICANT: Franzen, Henry
 APPLICANT: Chen, Xian
 APPLICANT: Evans, Sean
 APPLICANT: Shaked, Ze'ev
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 197
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 STATE STREET SUITE 510
 CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US 07/881,396
 FILING DATE: 08-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207
 REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 162:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 50.6%; Score 122; DB 6; Length 61;
 Best Local Similarity 93.8%; Pred. No. 5.99e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
 ||||||| |||||||
 QY 17 GCHGSEPCNIHRGKPF 32

RESULT 13

ID US-08-227-772-161 STANDARD; PRT; 84 AA.
 XX
 AC
 DT
 XX
 XX
 DE
 XX
 XX

Sequence 161, Application US/08227772

Sequence 161, Application US/08227772
 GENERAL INFORMATION:
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Rogers, Bruce
 APPLICANT: Franzen, Henry
 APPLICANT: Chen, Xian
 APPLICANT: Evans, Sean
 APPLICANT: Shaked, Ze'ev
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 207
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
 STREET: 610 LINCOLN STREET
 CITY: WALTHAM
 STATE: MA
 COUNTRY: USA
 ZIP: 02154

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/445,307A
 FILING DATE: 07 June 1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CRAIG, ANNE I.
 REGISTRATION NUMBER: 32,976
 REFERENCE/DOCKET NUMBER: 017.6US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 466-6000
 TELEFAX: (617) 466-6040
 INFORMATION FOR SEQ ID NO: 162:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 50.6%; Score 122; DB 8; Length 61;
 Best Local Similarity 93.8%; Pred. No. 5.99e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
 ||||||| |||||||
 QY 17 GCHGSEPCNIHRGKPF 32

RESULT 14

ID US-08-227-772-161 STANDARD; PRT; 84 AA.
 XX
 AC
 DT
 XX
 XX
 DE
 XX
 XX

Sequence 161, Application US/08227772

Sequence 161, Application US/08227772
 GENERAL INFORMATION:
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Rogers, Bruce
 APPLICANT: Franzen, Henry
 APPLICANT: Chen, Xian
 APPLICANT: Evans, Sean
 APPLICANT: Shaked, Ze'ev
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 207
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
 STREET: 610 LINCOLN STREET
 CITY: WALTHAM
 STATE: MA
 COUNTRY: USA
 ZIP: 02154

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/445,307A
 FILING DATE: 07 June 1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CRAIG, ANNE I.
 REGISTRATION NUMBER: 32,976
 REFERENCE/DOCKET NUMBER: 017.6US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 466-6000
 TELEFAX: (617) 466-6040
 INFORMATION FOR SEQ ID NO: 162:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 SEQUENCE 61 AA; 6669 MW; 17328 CN;

Query Match 50.6%; Score 122; DB 8; Length 61;
 Best Local Similarity 93.8%; Pred. No. 5.99e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
 ||||||| |||||||
 QY 17 GCHGSEPCNIHRGKPF 32

```
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 197
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;

Query Match 50.68; Score 122; DB 6; Length 84;
Best Local Similarity 93.8%; Pred. No. 5.99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||| |||||

RESULT 15
ID US-08-445-307A-161 STANDARD; PRT; 84 AA.
AC xxxxxx
XX
XX
XX
XX
XX
XX
XX
Sequence 161, Application US/08445307A
Sequence 161, Application US/08445307A
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
```

```
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;

Query Match 50.6%; Score 122; DB 8; Length 84;
Best Local Similarity 93.8%; Pred. No. 5.99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||| |||||
```

Search completed: Sun Apr 2 18:10:48 2000
Job time : 14 secs.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2000, 01:04:05 ; Search time 624.95 Seconds
(without alignments)
-1008.099 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagccgttctaagcgccgcg 175

Scoring table:
Gapop 10.0 , Gapext 1.0

Searched: 5086966 seqs, -1800032204 residues

Total number of hits satisfying chosen parameters: 10173922

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

- 1: /cgn2_6/ptodata/2/pna/PCTUS9_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US081A_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US081B_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US081C_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US081D_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US082A_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US082B_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US082C_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US083A_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US083B_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US084A_COMB.seq:*
- 14: /cgn2_6/ptodata/2/pna/US084B_COMB.seq:*
- 15: /cgn2_6/ptodata/2/pna/US084C_COMB.seq:*
- 16: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
- 17: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
- 18: /cgn2_6/ptodata/2/pna/US087A_COMB.seq:*
- 19: /cgn2_6/ptodata/2/pna/US087B_COMB.seq:*
- 20: /cgn2_6/ptodata/2/pna/US087C_COMB.seq:*
- 21: /cgn2_6/ptodata/2/pna/US088A_COMB.seq:*
- 22: /cgn2_6/ptodata/2/pna/US088B_COMB.seq:*
- 23: /cgn2_6/ptodata/2/pna/US088C_COMB.seq:*
- 24: /cgn2_6/ptodata/2/pna/US088D_COMB.seq:*
- 25: /cgn2_6/ptodata/2/pna/US089A_COMB.seq:*
- 26: /cgn2_6/ptodata/2/pna/US089B_COMB.seq:*
- 27: /cgn2_6/ptodata/2/pna/US089C_COMB.seq:*
- 28: /cgn2_6/ptodata/2/pna/US089D_COMB.seq:*
- 29: /cgn2_6/ptodata/2/pna/US089E_COMB.seq:*
- 30: /cgn2_6/ptodata/2/pna/US089F_COMB.seq:*
- 31: /cgn2_6/ptodata/2/pna/US089G_COMB.seq:*
- 32: /cgn2_6/ptodata/2/pna/US089H_COMB.seq:*
- 33: /cgn2_6/ptodata/2/pna/US089I_COMB.seq:*
- 34: /cgn2_6/ptodata/2/pna/US090A_COMB.seq:*
- 35: /cgn2_6/ptodata/2/pna/US090B_COMB.seq:*
- 36: /cgn2_6/ptodata/2/pna/US090C_COMB.seq:*
- 37: /cgn2_6/ptodata/2/pna/US090D_COMB.seq:*
- 38: /cgn2_6/ptodata/2/pna/US090E_COMB.seq:*
- 39: /cgn2_6/ptodata/2/pna/US091A_COMB.seq:*
- 40: /cgn2_6/ptodata/2/pna/US091B_COMB.seq:*
- 41: /cgn2_6/ptodata/2/pna/US091C_COMB.seq:*
- 42: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
- 43: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
- 44: /cgn2_6/ptodata/2/pna/US092C_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US092D_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US092E_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US093C_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US093D_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US093E_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US093F_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US094A_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US094B_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US094C_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US094D_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US094E_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US094F_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US095E_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US095F_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US096F_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US097D_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US097E_COMB.seq:*

76: /cgn2_6/ptodata/2/pna/US097F_COMB.seq:*

77: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

78: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

79: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

80: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*

81: /cgn2_6/ptodata/2/pna/US098E_COMB.seq:*

82: /cgn2_6/ptodata/2/pna/US098F_COMB.seq:*

83: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

84: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

85: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

86: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*

87: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*

88: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*

89: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*

90: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*

91: /cgn2_6/ptodata/2/pna/US100C_COMB.seq:*

92: /cgn2_6/ptodata/2/pna/US100D_COMB.seq:*

93: /cgn2_6/ptodata/2/pna/US100E_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	175	100.0	175	51	US-09-362-731-6
2	101	57.7	105	51	US-09-362-731-12
3	100	57.1	103	51	US-09-362-731-11
4	93	53.1	103	51	US-09-362-731-9
5	48.2	27.5	588	2	US-07-881-396A-3
6	48.2	27.5	588	8	US-08-227-772-3
7	48.2	27.5	588	14	US-08-445-307A-3
8	48.2	27.5	588	15	US-08-484-296-3
9	48.2	27.5	591	15	US-08-460-040-5
10	43.4	24.8	390	27	US-08-930-264-15
11	42.6	24.3	390	27	US-08-930-264-19
12	41.8	23.9	390	27	US-08-930-264-1
13	41.8	23.9	390	27	US-08-930-264-3
14	41.8	23.9	390	27	US-08-930-264-5
					Sequence 6, Appl
					Sequence 12, Appl
					Sequence 11, Appl
					Sequence 9, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 15, Appl
					Sequence 19, Appl
					Sequence 1, Appl
					Sequence 3, Appl
					Sequence 5, Appl

```
15 41.8 23.9 390 27 US-08-930-264-7 Sequence 7, Appli
16 41.8 23.9 390 27 US-08-930-264-9 Sequence 9, Appli
17 41.8 23.9 390 27 US-08-930-264-11 Sequence 11, Appli
18 41.8 23.9 390 27 US-08-930-264-13 Sequence 13, Appli
19 41.8 23.9 390 27 US-08-930-264-21 Sequence 21, Appli
20 41.8 23.9 390 27 US-08-930-264-23 Sequence 23, Appli
21 41.8 23.9 390 27 US-08-930-264-25 Sequence 25, Appli
22 41.8 23.9 390 27 US-08-930-264-27 Sequence 27, Appli
23 41.8 23.9 390 27 US-08-930-264-29 Sequence 29, Appli
24 41.8 23.9 390 27 US-08-930-264-31 Sequence 31, Appli
25 41.8 23.9 390 27 US-08-930-264-33 Sequence 33, Appli
26 41.8 23.9 390 27 US-08-930-264-35 Sequence 35, Appli
27 41.8 23.9 390 27 US-08-930-264-37 Sequence 37, Appli
28 41.8 23.9 390 27 US-08-930-264-39 Sequence 39, Appli
29 41.8 23.9 390 27 US-08-930-264-41 Sequence 41, Appli
30 41.8 23.9 390 27 US-08-930-264-43 Sequence 43, Appli
31 41.8 23.9 390 27 US-08-930-264-45 Sequence 45, Appli
32 41.8 23.9 390 27 US-08-930-264-47 Sequence 47, Appli
33 41.8 23.9 390 27 US-08-930-264-49 Sequence 49, Appli
34 41.8 23.9 390 27 US-08-930-264-51 Sequence 51, Appli
35 41.8 23.9 390 27 US-08-930-264-53 Sequence 53, Appli
36 41.8 23.9 390 27 US-08-930-264-55 Sequence 55, Appli
37 41.8 23.9 390 27 US-08-930-264-57 Sequence 57, Appli
38 41.8 23.9 390 27 US-08-930-264-59 Sequence 59, Appli
39 41.8 23.9 390 27 US-08-930-264-61 Sequence 61, Appli
40 41.8 23.9 390 27 US-08-930-264-63 Sequence 63, Appli
41 41.8 23.9 390 27 US-08-930-264-65 Sequence 65, Appli
42 41.8 23.9 390 27 US-08-930-264-67 Sequence 67, Appli
43 41.8 23.9 390 27 US-08-930-264-69 Sequence 69, Appli
44 41.8 23.9 390 27 US-08-930-264-71 Sequence 71, Appli
45 41.8 23.9 390 27 US-08-930-264-73 Sequence 73, Appli
```

ALIGNMENTS

```
RESULT 1
US-09-362-731-6
; Sequence 6, Application US/09362731
; GENERAL INFORMATION:
; APPLICANT: U.C.B. S.A.
; TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
; FILE REFERENCE: P. UCB.09/WO
; CURRENT APPLICATION NUMBER: US/09/362,731
; CURRENT FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 175
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-362-731-6
```

```
Query Match 100.0%; Score 175; DB 51; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.1e-49;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattccaccatggtcagatataaaagcaaatcttaatttaggtataactgaac 60
|||||
Db 1 gaattccaccatggtcagatataaaagcaaatcttaatttaggtataactgaac 60
|||||

QY 61 taggaggttgcctggttcagaccatgtatcatcgtggttaaacattcggcggtt 120
|||||
Db 61 taggaggttgcctggttccagaccatgtatcatcgtggttaaacattcggcggtt 120
|||||

QY 121 gtacaggaagtgccttgcattatcacagaggaagccgttcttaagcgccgc 175
|||||
Db 121 gtacaggaagtgccttgcattatcacagaggaagccgttcttaagcgccgc 175
|||||
```

```
RESULT 2
US-09-362-731-12/c
; Sequence 12, Application US/09362731
; GENERAL INFORMATION:
; APPLICANT: U.C.B. S.A.
; TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
; FILE REFERENCE: P. UCB.09/WO
; CURRENT APPLICATION NUMBER: US/09/362,731
; CURRENT FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 105
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-362-731-12
```

```
Query Match 57.7%; Score 101; DB 51; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.7e-24;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 ggttcagaaccatgtatcatcgtggttaaacattcggcggttgcaggaagtga 134
|||||
Db 105 GGTTCAGAACCATGTATCATTCATCGTGGTAAACCATTCGCGGTTGTACGGGAAGTGAG 46
|||||

QY 135 ccttgattatcacagaggaagccgttcttaagcgccgc 175
|||||
Db 45 CCTTGATTATACAGAGGAAGCCGTTCTAAGCGGCCGC 5
|||||
```

```
RESULT 3
US-09-362-731-11
; Sequence 11, Application US/09362731
; GENERAL INFORMATION:
; APPLICANT: U.C.B. S.A.
; TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
; FILE REFERENCE: P. UCB.09/WO
; CURRENT APPLICATION NUMBER: US/09/362,731
; CURRENT FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 103
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-362-731-11
```

```
Query Match 57.1%; Score 100; DB 51; Length 103;
Best Local Similarity 100.0%; Pred. No. 5.8e-24;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattccaccatggtcagatataaaagcaaatcttaatttaggtataactgaac 60
|||||
Db 4 gaattccaccatggtcagatataaaagcaaatcttaatttaggtataactgaac 63
|||||

QY 61 taggaggttgcctggttcagaccatgtatcatcgc 100
|||||
Db 64 taggaggttgcctggttcagaccatgtatcatcgc 103
|||||
```

```
RESULT 4
US-09-362-731-9
; Sequence 9, Application US/09362731
```

GENERAL INFORMATION:
; APPLICANT: U.C.B. S.A.
; TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
; FILE REFERENCE: P. UCB.09/WO
; CURRENT APPLICATION NUMBER: US/09/362,731
; CURRENT FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 103
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-362-731-9

Query Match 53.1% Score 93; DB 51; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.4e-21;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 gttgcatgttgcagaaacatgtatcatcattcgtgtaaacattcggcggtgtgtcacg 126
|||||
Db 11 gttgcatgttgcagaaacatgtatcattcgtgtaaacattcggcggtgtgtcacg 70
|||||
QY 127 gaagtgccttgcattatatacacagaggaagc 159
|||||
Db 71 gaagtgccttgcattatatacacagaggaagc 103
|||||

RESULT 5
US-07-881-396A-3
; Sequence 3, Application US/07881396A
; GENERAL INFORMATION:
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Garman, Richard
; APPLICANT: Greenstein, Julia
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/881,396A
; FILING DATE: 19920508
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 588 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS

LOCATION: 69...509
US-07-881-396A-3
Query Match 27.5% Score 48.2; DB 2; Length 588;
Best Local Similarity 94.3%; Pred. No. 5.2e-06;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 61 taggaggttgcctggttcagaaacatgtatcattcgtgtaaacattc 113
|||||
Db 172 TACCAGGATGCCATGTTTCAGAACCATGTCATTCATTCATCGTGGTAAACCATTC 224
|||||

RESULT 6
US-08-227-772-3
; Sequence 3, Application US/08227772
; GENERAL INFORMATION:
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Garman, Richard
; APPLICANT: Greenstein, Julia
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,772
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/881,396
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69...509
US-08-227-772-3

Query Match 27.5% Score 48.2; DB 8; Length 588;
Best Local Similarity 94.3%; Pred. No. 5.2e-06;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 61 taggaggttgcctggttcagaaacatgtatcattcgtgtaaacattc 113
|||||
Db 172 TACCAGGATGCCATGTTTCAGAACCATGTCATTCATTCATCGTGGTAAACCATTC 224
|||||

RESULT 7
US-08-445-307A-3

STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,040
FILING DATE: 2-JUNE-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/162,722

```

RESULT 10
US-08-930-264-15
; Sequence 15, Application US/08930264
; GENERAL INFORMATION:
; APPLICANT: NISHIYAMA, Chiharu
; APPLICANT: YUUKI, Toshifumi
; APPLICANT: OKUMURA, Yasushi
; TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
; TITLE OF INVENTION: PRODUCTION THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,264
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-93236
; FILING DATE: 28-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00791
; FILING DATE: 27-MAR-1996
; ATTORNEY/AGENT INFORMATION:

```

RESULT 11
US-08-930-264-19
US-08-930-264-19 Application US/08930264
Sequence 19, Application US/08930264
GENERAL INFORMATION:
APPLICANT: NISHIYAMA, Chiharu
APPLICANT: YUKI, Yoshifumi
APPLICANT: OKUMURA, Toshiaki
TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
TITLE OF INVENTION: PRODUCTION THEREOF
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,264
FILING DATE: 26-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-93236
FILING DATE: 28-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/00791
FILING DATE: 27-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: NISHIYAMA-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 390 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE:	CDS	NAME/KEY:	CDS	LOCATION:	1..387	US-08-930-264-19
Query Match	24.3%	Score 42.6;	DB 27;	Length 390;		
Best Local Similarity	91.8%	Pred. No. 0.00035;				
Matches	45;	Conservative	0;	Mismatches	4;	Indels 0; Gaps 0;
QY	65	aggttgccatggttcagaaacattatcatcattcatcggtgtaaacattc	113			
DB	57	AGGTGCGCATGTTCTGATCCATCATCCATCGTGGTAAACCATTC	105			
RESULT 12						
US-08-930-264-1						
Sequence 1, Application US/08930264						
GENERAL INFORMATION:						
APPLICANT: NISHIYAMA, Chiharu						
APPLICANT: YUUKI, Toshifumi						
APPLICANT: OKUMURA, Yasushi						
TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE						
PRODUCTION THEREOF						
NUMBER OF SEQUENCES: 40						
CORRESPONDENCE ADDRESS:						
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.						
STREET: 419 Seventh Street, N.W., Suite 300						
CITY: Washington						
STATE: D.C.						
COUNTRY: USA						
ZIP: 20004						
COMPUTER READABLE FORM:						
MEDIUM TYPE: Floppy disk						
COMPUTER: IBM PC compatible						
OPERATING SYSTEM: PC-DOS/MS-DOS						
SOFTWARE: Patent In Release #1.0, Version #1.30						
CURRENT APPLICATION DATA:						
APPLICATION NUMBER: US/08/930,264						
FILING DATE: 26-SEP-1997						
CLASSIFICATION: 536						
PRIOR APPLICATION DATA:						
APPLICATION NUMBER: JP 7-93236						
FILING DATE: 28-MAR-1995						
ATTORNEY/AGENT INFORMATION:						
NAME: BROWDY, Roger L.						
REGISTRATION NUMBER: 25,618						
REFERENCE/DOCKET NUMBER: NISHIYAMA-2						
TELECOMMUNICATION INFORMATION:						
TELEPHONE: 202-628-5197						
TELEFAX: 202-737-3528						
INFORMATION FOR SEQ ID NO: 1:						
SEQUENCE CHARACTERISTICS:						
LENGTH: 390 base pairs						
TYPE: nucleic acid						
STRANDEDNESS: single						
TOPOLOGY: linear						
MOLECULE TYPE: CDNA						
FEATURE:						
NAME/KEY: CDS						
LOCATION: 1..387						
US-08-930-264-1						
Query Match	23.9%	Score 41.8;	DB 27;	Length 390;		
Best Local Similarity	86.8%	Pred. No. 0.00066;				
Matches	46;	Conservative	0;	Mismatches	7;	Indels 0; Gaps 0;
QY	61	taggaggttgccatggttcagaaacattatcatcattcatcggtgtaaacattc	113			
DB	53	TCGATGGTTCATGCTGATCCATCATCCATCGTGGTAAACCATTC	105			
RESULT 14						
US-08-930-264-5						
Sequence 5, Application US/08930264						
GENERAL INFORMATION:						
APPLICANT: NISHIYAMA, Chiharu						
APPLICANT: YUUKI, Toshifumi						
APPLICANT: OKUMURA, Yasushi						
TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE						
PRODUCTION THEREOF						
NUMBER OF SEQUENCES: 40						
CORRESPONDENCE ADDRESS:						
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.						
STREET: 419 Seventh Street, N.W., Suite 300						
CITY: Washington						
STATE: D.C.						
COUNTRY: USA						
ZIP: 20004						

;; TITLE OF INVENTION: PRODUCTION THEREOF
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;; STREET: 419 Seventh Street, N.W., Suite 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/930,264
;; FILING DATE: 26-SEP-1997
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 7-93236
;; FILING DATE: 28-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP96/00791
;; FILING DATE: 27-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWDY, Roger L.
;; REGISTRATION NUMBER: 25,618
;; REFERENCE/DOCKET NUMBER: NISHIYAMA-2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-628-5197
;; TELEFAX: 202-737-3528
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 390 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..387
US-08-930-264-5

Query Match 23.9%; Score 41.8; DB 27; Length 390;
Best Local Similarity 86.8%; Pred. No. 0.00066;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 61 taggaggttgcacatgttcagaaccatgtatcatcattcgtgtaaacattc 113
Db 53 TCGATGGTTGCCATGGTTCGTGATCCATCATCCATCGTGGTAAACCATTC 105

RESULT 15
US-08-930-264-7
;; Sequence 7, Application US/08930264
;; GENERAL INFORMATION:
;; APPLICANT: NISHIYAMA, Chiharu
;; APPLICANT: YUUKI, Toshifumi
;; APPLICANT: OKUMURA, Yasushi
;; TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
;; TITLE OF INVENTION: PRODUCTION THEREOF
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;; STREET: 419 Seventh Street, N.W., Suite 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/930,264
;; FILING DATE: 26-SEP-1997
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 7-93236
;; FILING DATE: 28-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP96/00791
;; FILING DATE: 27-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWDY, Roger L.
;; REGISTRATION NUMBER: 25,618
;; REFERENCE/DOCKET NUMBER: NISHIYAMA-2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-628-5197
;; TELEFAX: 202-737-3528
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 390 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..387
US-08-930-264-7

Query Match 23.9%; Score 41.8; DB 27; Length 390;
Best Local Similarity 86.8%; Pred. No. 0.00066;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 61 taggaggttgcacatgttcagaaccatgtatcatcattcgtgtaaacattc 113
Db 53 TCGATGGTTGCCATGGTTCGTGATCCATCATCCATCGTGGTAAACCATTC 105

Search completed: April 8, 2000, 03:15:31
Job time: 7886 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:43:21 ; Search time 37.87 seconds
(without alignments)
553.249 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagccgttctaagcggcgcg 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCUS9_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.2	27.5	588	1 US-07-945-288-3	Sequence 3, Appli
2	48.2	27.5	588	1 US-08-462-831-3	Sequence 3, Appli
3	48.2	27.5	588	2 US-08-461-809-3	Sequence 3, Appli
4	48.2	27.5	588	2 US-08-461-441-3	Sequence 3, Appli
5	48.2	27.5	588	2 US-08-482-142-3	Sequence 3, Appli
6	48.2	27.5	588	4 US-08-478-572-3	Sequence 3, Appli
7	48.2	27.5	588	6 PCT-US93-08518-3	Sequence 3, Appli
8	41.8	23.9	513	2 US-08-288-888-3	Sequence 3, Appli
9	41.8	23.9	513	3 US-08-910-075-3	Sequence 3, Appli
10	41.8	23.9	513	4 US-08-905-801A-3	Sequence 3, Appli
11	41.8	23.9	516	2 US-08-288-888-1	Sequence 1, Appli
12	41.8	23.9	517	3 US-08-910-075-1	Sequence 1, Appli
13	41.8	23.9	517	4 US-08-905-801A-1	Sequence 1, Appli
14	40.2	23.0	491	1 US-07-945-288-7	Sequence 7, Appli
15	40.2	23.0	491	1 US-08-462-831-7	Sequence 7, Appli
16	40.2	23.0	491	2 US-08-461-809-7	Sequence 7, Appli
17	40.2	23.0	491	2 US-08-461-441-7	Sequence 7, Appli
18	40.2	23.0	491	2 US-08-482-142-7	Sequence 7, Appli
19	40.2	23.0	491	4 US-08-478-572-7	Sequence 7, Appli
20	40.2	23.0	491	6 PCT-US93-08518-7	Sequence 7, Appli
21	29.2	16.7	1474	2 US-08-465-980-1	Sequence 1, Appli
22	29.2	16.7	1474	4 US-09-053-303-1	Sequence 1, Appli
23	29.2	16.7	1474	6 PCT-US95-07093-1	Sequence 1, Appli
24	28.2	16.1	1151	1 US-08-236-754-3	Sequence 1, Appli
25	28.2	16.1	3055	1 US-08-236-754-1	Sequence 3, Appli
26	27	15.4	2380	2 US-08-572-951-3	Sequence 3, Appli
27	27	15.4	3142	4 US-08-956-242-3	Sequence 3, Appli

C 28 26.8 15.3 5359 4 US-09-070-060-1 Sequence 1, Appli
C 29 26.6 15.2 1315 3 US-08-578-592-4 Sequence 4, Appli
30 2973 2 US-08-451-715A-7 Sequence 7, Appli
31 26.4 15.1 1626 2 US-08-604-333-1 Sequence 1, Appli
32 26.2 15.0 2951 1 US-08-413-118-104 Sequence 104, App
33 26.2 15.0 3000 1 US-08-220-151-1 Sequence 1, Appli
34 26.2 15.0 3000 1 US-08-413-118-1 Sequence 1, Appli
35 26 14.9 1503 1 US-08-307-499-41 Sequence 41, Appli
C 36 26 14.9 14176 1 US-08-307-499-1 Sequence 1, Appli
37 26 14.9 14176 1 US-08-307-499-14 Sequence 14, Appli
38 25.6 14.6 54 3 US-08-661-052-7 Sequence 7, Appli
C 39 25.2 14.4 1734 2 US-08-551-437-4 Sequence 4, Appli
40 25.2 14.4 5495 2 US-08-602-010A-1 Sequence 1, Appli
C 41 25.2 14.4 5495 2 US-08-602-010A-2 Sequence 2, Appli
42 25.2 14.4 5495 2 US-08-680-726A-1 Sequence 1, Appli
C 43 25.2 14.4 5495 2 US-08-680-726A-2 Sequence 2, Appli
C 44 25 14.3 2172 1 US-08-158-232-50 Sequence 50, Appli
45 25 14.3 2172 2 US-08-611-928-50 Sequence 50, Appli

ALIGNMENTS

RESULT 1
US-07-945-288-3
; Sequence 3, Application US/07945288
; Patent No. 5433948
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kow-Yan
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,288
FILING DATE: 19920910
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 69..509
US-07-945-288-3

COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,441
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,388
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 69..509
US-08-461-441-3

Query Match 27.5%; Score 48.2; DB 2; Length 588;
Best Local Similarity 94.3%; Pred. No. 2.5e-07;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 61 taggaggttgcattggttcagaaaccatgtatcattcattcgtggtgtaaacattc 113
Db 172 TACCAGGATGCCATGTTCCAGAACCAATGATCATTCATCGTGGAACCAATTC 224

RESULT 5
US-08-482-142-3
Sequence 3, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,142
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,605
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 69..509
US-08-482-142-3

Query Match 27.5%; Score 48.2; DB 2; Length 588;
Best Local Similarity 94.3%; Pred. No. 2.5e-07;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 61 taggaggttgcattggttcagaaaccatgtatcattcattcgtggtgtaaacattc 113
Db 172 TACCAGGATGCCATGTTCCAGAACCAATGATCATTCATCGTGGAACCAATTC 224

RESULT 6
US-08-478-572-3
Sequence 3, Application US/08478572
Patent No. 5988526
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,572
FILING DATE: 07-June-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE:


```

1  TITLE OF INVENTION:  A PROTEIN ALLERGEN OF Derf II AND
2  TITLE OF INVENTION:  COMPOSITION (as amended)
3  NUMBER OF SEQUENCES:  4
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  BROWDY AND NEIMARK
6  STREET:  419 Seventh Street, N.W., Suite 300
7  CITY:  Washington
8  STATE:  D.C.
9  COUNTRY:  USA
10 ZIP:  20004
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE:  Floppy disk
13 COMPUTER:  IBM PC compatible
14 OPERATING SYSTEM:  PC-DOS/MS-DOS
15 SOFTWARE:  PatentIn Release #1.0, Version #1.30
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER:  US/08/905,801A
18 FILING DATE:
19 CLASSIFICATION:  536
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER:  JP 2-50848
22 FILING DATE:  03-MAR-1990
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER:  US 07/658,596
25 FILING DATE:  21-FEB-1991
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER:  US 08/288,888
28 FILING DATE:  10-AUG-1994
29 ATTORNEY/AGENT INFORMATION:
30 NAME:  BROWDY, Roger L.
31 REGISTRATION NUMBER:  25,618
32 REFERENCE/DOCKET NUMBER:  YUUKI-1C
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE:  202-628-5197
35 TELEFAX:  202-737-3528
36 INFORMATION FOR SEQ ID NO:  3:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH:  513 base pairs
39 TYPE:  nucleic acid
40 STRANDEDNESS:  single
41 TOPOLOGY:  linear
42 MOLECULE TYPE:  cdna
43 FEATURE:
44 NAME/KEY:  CDS
45 LOCATION:  1..426
46 US-08-905-801A-3
47
48 Query Match 23.9% Score 41.8; DB 4; Length 513;
49 Best Local Similarity 86.8%; Pred. No. 3.2e-05;
50 Matches 46; Conservative 0; Mismatches 7; Indels 0;
51
52 Qy 61 tagagagttgccatggttcagaaaccatgtatcatctcgtgtgtaaacattc 113
53 + | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
54 Db 92 TCGATGTTGCCATGGTTCTCATCCATCATCCATCGTGGTAAACCAATC 144
55
56 RESULT 11
57 US-08-288-888-1
58 Sequence 1, Application US/08288888
59 Patent No. 5798099
60 GENERAL INFORMATION:
61 APPLICANT:  YUUKI, Toshifumi
62 APPLICANT:  OKUMURA, yasushi
63 APPLICANT:  YAMAKAWA, Hiroshi
64 TITLE OF INVENTION:  DNA ENCODING A MAJOR MITE ALLERGEN AND
65 TITLE OF INVENTION:  THE USE THEREOF
66 NUMBER OF SEQUENCES:  4
67 CORRESPONDENCE ADDRESS:
68 ADDRESSEE:  BROWDY AND NEIMARK
69 STREET:  419 SEVENTH ST., N.W., SUITE 300
70 CITY:  WASHINGTON
71 STATE:  DC

```


ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: YUUKI-1C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 517 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..426
US-08-905-801A-1

Query Match 23.9%; Score 41.8; DB 4; Length 517;
Best Local Similarity 86.8%; Pred. No. 3.2e-05;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 61 taggaggttgcattggttcagaaaccatgtatcatcattcatcattcgtggttaaacattc 113
Db 92 TCGATGGTGGCATTGTTCTGATCCATCATCATCCATCGTGGTAAACATTTC 144

RESULT 14
US-07-945-288-7
Sequence 7, Application US/07945288
Patent No. 5433948
GENERAL INFORMATION:
APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kow-Yan
TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,288
FILING DATE: 19920910
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 491 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..390
US-07-945-288-7

Query Match 23.0%; Score 40.2; DB 1; Length 491;
Best Local Similarity 84.9%; Pred. No. 0.00011;
Matches 45; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 61 taggaggttgcattggttcagaaaccatgtatcatcattcatcattcgtggttaaacattc 113
Db 53 TCGATGGTGGCATTGTTCTGATCCATCATCATCCATCGTGGTAAACATTTC 105

RESULT 15
US-08-462-831-7
Sequence 7, Application US/08462831
Patent No. 5552142
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
TITLE OF INVENTION: DERMATOPHAGOIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,831
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 491 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..390
US-08-462-831-7

Query Match 23.0%; Score 40.2; DB 1; Length 491;
Best Local Similarity 84.9%; Pred. No. 0.00011;
Matches 45; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 61 taggaggttgcattggttcagaaaccatgtatcatcattcatcattcgtggttaaacattc 113

Db 53 TCGATGGTGGCCATGGTTCTGTGATCCATGCATATCCATCCATCGTGGTAACCATTC 105

Search completed: April 8, 2000, 02:55:17
Job time: 11516 sec